Information Theory in Property Testing and Monotonicity Testing in Higher Dimension*

Nir Ailon[1] and Bernard Chazelle[1]

Department of Computer Science, Princeton University, Princeton NJ, USA
{nailon,chazelle}@cs.princeton.edu

Abstract. In general property testing, we are given oracle access to a function f, and we wish to randomly test if the function satisfies a given property P, or it is ε-far from having that property. In a more general setting, the domain on which the function is defined is equipped with a probability distribution, which assigns different weight to different elements in the distance function. This paper relates the complexity of testing the monotonicity of a function over the d-dimensional cube to the Shannon entropy of the underlying distribution. We provide an improved upper bound on the property tester query complexity and we finetune the exponential dependence on the dimension d.

1 Introduction

In general property testing [4,7,9,13], we are given oracle access to a function f, and we wish to randomly test if the function satisfies a given property P, or it is ε-far from having that property. By ε-far we mean, that any function g that has the property P differs from f in at least ε-fraction places. We allow the property tester to err with at most constant probability, say $1/3$ (in this paper we assume only one-sided error). In many interesting cases, this relaxation allows the tester to query only a sublinear portion of the input f, which is crucial when the input is a giant dataset.

The *query complexity* of the property is the minimal number of f-queries performed by a tester for that property (although the classical "number of operations" quantity can be considered too). A query to a function can be viewed as a quantity of information, which gives rise to the relation between property testing and information complexity [4], which will be made more precise in what follows.

An interesting ramification of property testing problems [4,5,10] generalizes the definition of distance between two functions: Instead of defining the distance between f and g as the fractional size of the set $\{x \mid f(x) \neq g(x)\}$, we attach a probability distribution $\mathcal{D}$ to the function domain, and define

$$\operatorname{dist}(f, g) = \Pr(\{x \mid f(x) \neq g(x)\}).$$

The "old" definition reduces to the case $\mathcal{D} = \mathcal{U}$ (the uniform distribution). This definition allows assignment of importance weights to domain points. It also allows property testers to deal with functions defined on infinite domains, though it may be necessary to assume additional structure (for example, measurability of f). Such functions arise when dealing with natural phenomena, like the temperature as a function of location and time. Of course in these cases we couldn't read the entire input even if we had unlimited resources.

The distribution should not be considered as part of the problem, but rather as a parameter of the problem. Fischer [4] distinguishes between the case where $\mathcal{D}$ is known to the tester, and the case where it is not known. The latter is known as the "distribution-free" case [10]. In the distribution-free case, the property tester is allowed to sample from the distributiun (but it does not know the probabilities). The main techniques developed in this work will be used for the distribution-known case, but we will also show an application to the distribution-free case.

The following question motivated the results in this paper: what happens when the distribution $\mathcal{D}$ is uniform on a strict subset S of the domain, and zero outside S? Intuitively, the "effective" domain is smaller, and therefore testing the property should be simpler. For general distributions, a natural measure of the "size" of the effective domain is the Shannon entropy H of $\mathcal{D}$. In this paper we show a connection between the quantity H and the query complexity, which further supports the connection between property testing and information theory.

* This work was supported in part by NSF grant CCR-0306283 and ARO Grant DAAH04-96-1-0181.

One interesting, well-studied property is monotonicity [2–4, 6, 8, 10–12]. A real function f over a poset $\mathcal{P}$ is monotone if any $x, y \in \mathcal{P}$ such that $x \leq y$ satisfy $f(x) \leq f(y)$. In this paper we assume that $\mathcal{P}$ is the d-dimensional cube $[n]^d$, with the order:

$$(x_1, \ldots, x_d) \leq (y_1, \ldots, y_d) \text{ if } x_i \leq y_i \text{ for all } i = 1, \ldots, d.$$

Halevy and Kushilevitz [10] describe a property tester with query complexity $O(\frac{2^d \log^d n}{\varepsilon})$ in the distribution-free case. In [11] they show a property tester with query complexity $O(\frac{d4^d \log n}{\varepsilon})$, for the special case of known uniform distribution ($\mathcal{D} = \mathcal{U}$). If d is fixed, this result improves a result by Dodis et al. [2], who describe a property tester with query complexity $O(\frac{d^2 \log^2 n}{\varepsilon})$ (For large d, n must be doubly-exponential in d for Halevy-Kushilevitz's result to be better than that of Dodis et al.).

The main result of our paper is as follows:

Theorem 1. *Let $\mathcal{D}$ be a (known) distribution on $[n]^d$ with independent marginal distributions (in other words, $\mathcal{D}$ is a product $\mathcal{D}_1 \times \cdots \times \mathcal{D}_d$ of distributions $\mathcal{D}_i$ on $[n]$). Let H be the Shannon entropy of $\mathcal{D}$. Then there exists a property tester for functions over $([n]^d, \mathcal{D})$ with expected query complexity $O(\frac{2^d H}{\varepsilon})$.*

In the special case $\mathcal{D} = \mathcal{U}$, this theorem improves Halevy and Kushilevitz's result by replacing the 4^d with 2^d (because then $H = d \log n$). It also generalizes previous work to any product distribution and gives a first evidence of the connection between property testing and the Shannon entropy of the underlying distribution.

Although this paper discusses mainly the *known* distribution case, the techniques developed here can be used to show the following:

Theorem 2. *Let $\mathcal{D}$ be an (unknown) distribution on $[n]^d$ with independent marginal distributions. Then there exists a property tester for functions over $([n]^d, \mathcal{D})$ with query complexity $O(\frac{d2^d \log n}{\varepsilon})$.*

Note that although Theorem 2 assumes that the distribution $\mathcal{D}$ is unknown, it will in fact be implicitly assumed by the property tester that $\mathcal{D}$ is a product of d marginal distributions. This is a relaxation of the notion of distribution-free property testing: the distribution is assumed to belong to some big family of distributions. This improves Halevy and Kushilevitz's $O(\frac{\log^d n 2^d}{\varepsilon})$ property tester [10] for this relaxed version (in their result, however, nothing is assumed about the distribution $\mathcal{D}$).

The rest of the paper is organized as follows: Section 2 starts with preliminaries and definitions, Section 3 proves Theorem 1 for the case $([n], \mathcal{D})$, Section 4 proves Theorem 1 for the case $([n]^d, \mathcal{U})$, and Section 5 completes the proof of Theorem 1. In Section 6 we prove Theorem 2. Section 7 discusses future work and open problems.

2 Preliminaries

Let f be a real valued function on the domain $[n]^d$, with a probability distribution $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_d$. Assume that D_i assigns probability p_j^i to $j \in [n]$, and therefore $\mathcal{D}$ assigns probability $\prod_{k=1}^d p_{i_k}^k$ to $(i_1, i_2, \ldots, i_d)$.

Definition 1. *The distance of f from monotonicity, denoted by ε, is $\min \Pr_{\mathcal{D}}(\{f \neq g\})$, where the minimum is over all monotone functions g.*

We will also use the notion of the axis-parallel ("projected") order.

Definition 2. *The i-th axis-parallel order $\leq_i$ on $[n]^d$ is defined as*

$$(x_1, \ldots, x_d) \leq_i (y_1, \ldots, y_d) \textit{ if } x_i \leq y_i \textit{ and } x_j = y_j \textit{ for } j \neq i.$$

Definition 3. *The i-th axis-parallel distance of f to monotonicity, denoted by ε_i, is $\min \Pr_{\mathcal{D}}(\{f \neq g\})$, where the minimum is over all functions g that are monotone with respect to $\leq_i$.*

It is a simple observation that f is monotone on $[n]^d$ if and only if it is monotone with respect to $\leq_i$ for each $i = 1, \ldots, d$.

Definition 4. *An integer pair $\langle i,j\rangle$ (for $i,j \in [n]^d$, $i \le j$) is a* violating pair *if $f(i) > f(j)$. We say that "j is in violation with i" or "i is in violation with j" in this case.*

Although this work deals with the finite domain case, it will be useful in what follows to consider the continuous cube I^d, where $I = \{x \in \mathbb{R} \mid 0 \le x < 1\}$. The probability distribution is the Lebesgue measure, denoted by μ. The distance between two measurable functions $\alpha, \beta : I^d \to \mathbb{R}$ is $\mu(\{\alpha \neq \beta\})$ (the set $\{\alpha \neq \beta\}$ is measurable). The distance of α from monotonicity is $\inf \operatorname{dist}(\alpha, \beta)$ where the infimum is over all monotone functions β.

For $i = 1, \ldots, d$, consider the following sequence of subintervals covering I:

$$\Delta_1^i = [0, p_1^i), \Delta_2^i = [p_1^i, p_1^i + p_2^i), \ldots, \Delta_n^i = [1 - p_n^i, 1).$$

For a number $x \in I$, define $\operatorname{int}_i(x) = j$ if $x \in \Delta_j^i$, that is, x belongs to the j-th interval induced by $\mathcal{D}_i$. If $d = 1$ we omit the superscript and simply write Δ_j and $\operatorname{int}(x)$. It is obvious that if x is distributed uniformly in I, then $\operatorname{int}_i(x)$ is distributed according to $\mathcal{D}_i$.

For a given $f : [n]^d \to \mathbb{R}$, denote by $\tilde{f} : I^d \to \mathbb{R}$ the function

$$\tilde{f}(x_1, \ldots, x_d) = f(\operatorname{int}_1(x_1), \operatorname{int}_2(x_2), \ldots, \operatorname{int}_d(x_d)).$$

The function $\tilde{f}$ is constant on rectangles of the form $\Delta_{i_1}^1 \times \cdots \times \Delta_{i_d}^d$, for any $i_1, \ldots, i_d \in [n]$. Moreover, any function $\alpha : I^d \to \mathbb{R}$ which is constant on these rectangles can be viewed as a function over $[n]^d$. The following lemma formalizes an intuitive connection between $([n]^d, \mathcal{D})$ and $(I^d, \mathcal{U})$. The proof is postponed to Appendix A.

Lemma 1. *The distance $\tilde{\varepsilon}$ of $\tilde{f}$ from monotonicity in I^d (with respect to the Lebesgue measure) equals the distance ε of f from monotonicity in $[n]^d$ (with respect to $\mathcal{D}$). This is also true with respect to the axis-parallel orders $\le_i$.*

Finally, we give a precise definition of a property tester:

Definition 5. *An ε-property tester for monotonicity (or, ε-monotonicity tester) is a randomized algorithm that, given $f : [n]^d \to \mathbb{R}$, outputs "ACCEPT" with probability 1 if f is monotone, and "REJECT" with probability at least $2/3$ if f is ε-far from being monotone w.r.t. a fixed distribution $\mathcal{D}$. In the distribution-known case, the probabilities of $\mathcal{D}$ are known. In the distribution-free case they are unknown, but the property tester can sample from $\mathcal{D}$.*

Remark: If the distribution $\mathcal{D}$ has elements with probability 0, then a function can have distance 0 to monotonicity without being actually monotone. In our terminology, a "monotone function" is monotone in the traditional case (even elements with probability 0 must comply). Since we will always assume $\varepsilon > 0$, a function that is ε-far from being monotone is never a monotone function.

3 A property tester for $([n], \mathcal{D})$

The algorithm is a generalization of an algorithm presented in [10]. Let $f : [n] \to \mathbb{R}$ be the input function. We need a few definitions and lemmas.

Definition 6. *For a violating pair $\langle i,j\rangle$ we say that i is* active *if*

$$\Pr(\textit{ in violation with } i \mid [i+1, j]) \ge 1/2.$$

Similarly, j is active *if*

$$\Pr(\textit{ in violation with } j \mid [i, j-1]) \ge 1/2.$$

In other words, an active integer in a violating pair $\langle i,j\rangle$ is also in violation with an abundance of elements in the interval $[i,j]$.

Definition 7. *For a violating pair $\langle i,j\rangle$, we say that i is* strongly *active if it is active and $p_i \le Pr([i+1, j])$. Similarly, j is* strongly *active if it is active and $p_j \le Pr([i, j-1])$.*

Lemma 2. *If $\langle i, j \rangle$ is a violating pair, then either i is strongly active or j is strongly active.*

Proof. It is immediate that for any $i < k < j$, either $\langle i, k \rangle$ or $\langle k, j \rangle$ is a violating pair. So either i or j is in violation with at least half the weight of the integers $[i + 1, j - 1]$. This proves that either i or j is active. So assume i is active but *not* strongly active. This means that $p_i > Pr([i + 1, j])$. But this would imply that j is strongly active. Indeed, p_i is greater than half of $Pr([i, j - 1])$, and i is in violation with j, so j is active. But $p_j < p_i$ so j is strongly active. $\square$

Lemma 3. *Let J be the collection of strongly active integers from all violating pairs of f. Then $Pr(J) \geq \varepsilon$.*

Proof. Actually, any collection J of at least one integer from each violating pair has this property. Proof of this simple fact can be found in [10]. $\square$

To describe the algorithm, we need another piece of notation. For $x \in I$, let left(x) denote the left endpoint of the interval $\Delta_{\text{int}(x)}$, and similarly let right(x) denote its right endpoint.

The following algorithm is an ε-property tester for monotonicity of f, with expected query complexity $O(\frac{H+1}{\varepsilon})$. We show how to eliminate the added $1/\varepsilon$ shortly.

monotonicity-test $(f, \mathcal{D}, \varepsilon)$

```
1 repeat O(ε^{-1}) times
2     choose random x ∈ I
        set δ ← p_{int(x)}
3       set r ← right(x)
4       while r + δ ≤ 2
5           choose random y ∈_U [r, min{r + δ, 1}]
6           if f(int(x)) > f(int(y))
7             then output REJECT
            δ ← 2δ
        set δ ← p_{int(x)}
        set l ← left(x)
8       while l − δ ≥ −1
            choose random y ∈ [max{l − δ, 0}, x]
            if f(int(y)) > f(int(x))
              then output REJECT
            set δ ← 2δ
  output ACCEPT
```

We first calculate the expected running time of **monotonicity-test**. The number of iterations of the internal *while* loops (lines 4,8) is clearly at most $\log(2/p_{\text{int}(x)})$ (all the logarithms are taken in base 2 in this paper). Clearly,

$$\mathbf{E}_{x \in_{\mathcal{U}} I}[\log(2/p_{\text{int}(x)})] = \mathbf{E}_{i \in \mathcal{D}}[\log(2/p_i)] = H + 1.$$

We prove correctness of the algorithm. Obviously, if f is monotone then the algorithm returns "ACCEPT". Assume that f is ε-far from being monotone. By lemma 3, with probability at least ε, the random variable x chosen in line 2 satisfies $\text{int}(x) \in J$. This means that $i = \text{int}(x)$ is strongly active with respect to a violating pair $\langle i, j \rangle$ or $\langle j, i \rangle$. Assume the former case (a similar analysis can be done for the latter). So i is in violation with at least half the weight of $[i+1, j]$, and also $p_i \leq Pr([i+1, j])$. Consider the intervals $[r, r + p_i 2^t]$ for $t = 0, 1, 2, \ldots$ with r as in line 3. For some t, this interval "contains" the corresponding interval $[i + 1, j]$ (i.e. $\Delta_{i+1} \cup \cdots \cup \Delta_j$), but $p_i 2^t$ is at most twice $\Pr([i + 1, j])$. The latter by virtue of i being *strongly* active. For this t, with probability at least $1/2$ the y chosen in line 5 is in $[i + 1, j]$. In such a case, the probability of y being a witness of nonmonotonicity in lines 6-7 is at least $1/2$, by virtue of i being *active*. Summing up, we get that the probability of outputting "REJECT" in a single iteration of the loop in line 1 is at least $\varepsilon/4$. Repeating $O(\varepsilon^{-1})$ times gives a constant probability.

We note that the additive constant 1 in the query complexity can be eliminated using a simple technical observation. Indeed, notice that, for x chosen in line 2, if $p_{\mathrm{int}(x)} > 1/2$ then x cannot be strongly active by definition, and therefore that iteration can be aborted without any query. If $p_{\mathrm{int}(x)} \le 1/2$ then we can eliminate one iteration from the while loops by initializing $\delta = 2p_{\mathrm{int}(x)}$ instead of $\delta = p_{\mathrm{int}(x)}$ and by slightly decreasing the probability of success in each iteration of the *repeat* loop. This gets rid of the additive constant, and concludes the proof of Theorem 1 in the $([n], \mathcal{D})$ case.

4 A property tester for $([n]^d, \mathcal{U})$

Let $f : [n]^d \to \mathcal{U}$ denote the input function. For a dimension $j \in [d]$ and integers $i_1, \dots, \hat{i}_j, \dots, i_d \in [n]$, let $f^j_{i_1,\dots,\hat{i}_j,\dots,i_d}$ denote the one-dimensional function obtained by restricting f to the line

$$\{i_1\} \times \cdots \times \{i_{j-1}\} \times [n] \times \{i_{j+1}\} \times \cdots \times \{i_d\} .$$

highdim-mon-uniform-test (f, ε)

```
  repeat O(ε^{-1} d 2^d) times
1     choose random dimension j ∈ [d]
2     choose random i_1, ..., î_j, ..., i_d ∈ [n]
3     run one iteration of repeat loop of monotonicity-test(f^j_{i_1,...,î_j,...,i_d}, U, *)
  output ACCEPT
```

To prove that the above algorithm is an ε-monotonicity tester for f, we will need the following lemma. It is an improved version of a theorem from [11], with 2^d replacing the 4^d on the right hand side. Recall Definition 2 of ε_i.

Lemma 4.

$$\sum_{i=1}^{d} \varepsilon_i \ge \varepsilon / 2^{d+1}.$$

The correctness of **highdim-mon-uniform-test** is a simple consequence of Lemma 4. If f is monotone, then the algorithm returns "ACCEPT" with probability 1. So assume f is ε-far from monotonicity. By Lemma 4, the restricted one-dimensional function $f^j_{i_1,\dots,\hat{i}_j,\dots,i_d}$ chosen in line 3 has expected distance of at least $\gamma = \frac{1}{d} \sum \varepsilon_i \ge \frac{1}{d} \varepsilon / 2^{d+1}$ from monotonicity, in each iteration of the *repeat* loop. A single iteration of **monotonicity-test** has an expected success probability of $\Omega(\gamma)$ by the analysis of the previous section. Repeating $O(\varepsilon^{-1} d 2^d)$ times amplifies the probability of success to any fixed constant. As for the query complexity, line 3 makes $O(\log n)$ queries, which is the entropy of the uniform distribution on $[n]$. So the entire query complexity is $O(\varepsilon^{-1} 2^d d \log n) = O(\varepsilon^{-1} 2^d H)$, as required. It remains to prove Lemma 4:

Proof. For $i = 1, \dots, d$, let B_i denote a minimal subset of $[n]^d$ such that f can be changed on B_i to get a monotone function with respect to $\le_i$. So $|B_i| = n^d \varepsilon_i$. Let $B = \cup_{i=1}^d B_i$. So $|B| \le \sum \varepsilon_i [n]^d$. Let $\chi_B : [n]^d \to \{0, 1\}$ denote the characteristic function of B:

$$\chi_B(x) = \begin{cases} 1 & x \in B \\ 0 & \text{otherwise} \end{cases}$$

We define operators Ψ_L and Ψ_R on $\{0, 1\}$ functions over $[n]$ as follows:

$$(\Psi_L v)(i) = \begin{cases} 1 & \text{if there exists } j \in [1, i] \text{ s.t. } \sum_{k=j}^{i} v(k) \ge (i - j + 1)/2 \\ 0 & \text{otherwise} \end{cases}$$

$$(\Psi_R v)(i) = \begin{cases} 1 & \text{if there exists } j \in [i, n] \text{ s.t. } \sum_{k=i}^{j} v(k) \ge (j - i + 1)/2 \\ 0 & \text{otherwise} \end{cases}$$

Given a $\{0,1\}$-function over $[n]^d$, we define operators $\Psi_L^{(i)}$ (resp. $\Psi_R^{(i)}$) for $i = 1, \ldots, d$ by applying Ψ_R (resp. Ψ_L) independently on one-dimensional lines of the form

$$\{x_1\} \times \cdots \times \{x_{i-1}\} \times [n] \times \{x_{i+1}\} \times \cdots \times \{x_d\}.$$

Finally, for $i = 1, \ldots, d$ we define the functions $\varphi_L^{(i)}, \varphi_R^{(i)} : [n]^d \to \{0,1\}$ as follows:

$$\begin{aligned} \varphi_L^{(i)} &= \left(\Psi_L^{(i)} \circ \Psi_L^{(i+1)} \circ \cdots \circ \Psi_L^{(d)}\right) \chi_B \\ \varphi_R^{(i)} &= \left(\Psi_R^{(i)} \circ \Psi_R^{(i+1)} \circ \cdots \circ \Psi_R^{(d)}\right) \chi_B \end{aligned} \tag{1}$$

Note that $\varphi_L^{(i)} = \Psi_L^{(i)} \varphi_L^{(i+1)}$ and $\varphi_R^{(i)} = \Psi_R^{(i)} \varphi_R^{(i+1)}$.

We claim that outside the set $\{\varphi_L^{(1)} = 1\} \cup \{\varphi_R^{(1)} = 1\} \subseteq [n]^d$ the function f is monotone. Indeed, choose $x, y \in [n]^d$ such that $x \leq y$ and $\varphi_L^{(1)}(y) = \varphi_R^{(1)}(x) = 0$. We want to show that $f(x) \leq f(y)$.

Claim 3 *Any $b \in B$ satisfies $\varphi_L^{(i)}(b) = \varphi_R^{(i)}(b) = 1$ for $i = 1, \ldots, d$.*

By the above Claim, $x, y \notin B$. Now consider the two line segments:

$$\begin{aligned} S_R &= [x_1, y_1] \times \{x_2\} \times \cdots \times \{x_d\} \\ S_L &= [x_1, y_1] \times \{y_2\} \times \cdots \times \{y_d\}. \end{aligned}$$

By definition of $\Psi_R^{(1)}$ (resp. $\Psi_L^{(1)}$), the average value of $\varphi_R^{(2)}$ (resp. $\varphi_L^{(2)}$) on S_R (resp. S_L) is less than $1/2$. Therefore, there exists $z_1 \in [x_1, y_1]$ such that $\varphi_R^{(2)}(z_1, x_2, \ldots, x_d) + \varphi_L^{(2)}(z_1, y_2, \ldots, y_d) < 1$. Since these values are in $\{0,1\}$, we get that

$$\varphi_R^{(2)}(z_1, x_2, \ldots, x_d) = \varphi_L^{(2)}(z_1, y_2, \ldots, y_d) = 0. \tag{2}$$

Denote $x^{(1)} = (z_1, x_2, \ldots, x_d)$ and $y^{(1)} = (z_2, y_2, \ldots, y_d)$. By Claim 3 and (2), both $x^{(1)}$ and $y^{(1)}$ are outside B. Since $x \leq_1 x^{(1)}$ we get that $f(x) \leq f(x^{(1)})$. A similar argument shows that $f(y^{(1)}) \leq f(y)$.

We use an inductive argument, using the functions $\varphi_L^{(2)}$ and $\varphi_R^{(2)}$ to show that $f(x^{(1)}) \leq f(y^{(1)})$. The general inductive step generates points $x^{(i)} \leq y^{(i)}$ that agree in the first i coordinates, and such that $\varphi_R^{(i+1)}(x^{(i)}) = \varphi_L^{(i+1)}(y^{(i)}) = 0$ (consequently, $x^{(i)}, y^{(i)} \notin B$).

In the base step we will end up with $x^{(d-1)}$ and $y^{(d-1)}$ that differ in their last coordinate only. Therefore, they are $\leq_d$-comparable and $f(x^{(d-1)}) \leq f(y^{(d-1)})$ because $x^{(d-1)}, y^{(d-1)} \notin B$.

It remains to bound the size of the set $\{\varphi_L^{(1)} = 1\}$. A similar analysis can be applied to $\{\varphi_R^{(1)} = 1\}$. We claim that $|\{\varphi_L^{(1)} = 1\}| \leq |B| 2^d$. This is a simple consequence of the following lemma.

Lemma 5. *Let $v \in \{0,1\}^n$. Then the number of 1's in $\Psi_L v$ is at most twice the number of 1's in v. A similar result holds for Ψ_R.*

To prove this, imagine walking on the domain $[n]$ from 1 to n, and marking integers according to the following rule (assume on initialization that all domain points are unmarked and a counter is set to 0):

If the value of v on the current integer i is 1, then mark i. Also, in this case increase the counter by 1. If $v(i) = 0$ and the counter is > 0, then mark integer i and decrease the counter by 1. Otherwise do nothing.

It is obvious that the number of marked integers is at most twice the number of 1's in v. It is also not hard to show that $(\Psi_L v)(i) = 1$ only if i is marked. Indeed, if $(\Psi_L v)(i) = 1$, then for some $j \leq i$, vector v on integer segment $[j, i]$ has at least as many 1's as 0's. This implies that either $v(i) = 1$ or the counter at i is positive, therefore i is marked. This proves the lemma.

We conclude that the combined size of $\{\varphi_L^{(1)} = 1\}$ and $\{\varphi_R^{(1)} = 1\}$ is at most $|B| 2^{d+1}$. This means that f is monotone on a subset of $[n]^d$ of size at least $n^d - |B| 2^{d+1}$. It is a simple fact that any monotone function on a subset of $[n]^d$ can be completed to a monotone function on the entire domain (see Lemma 1 [6]). So the distance ε of f from monotonicity is at most $2^{d+1} \sum \varepsilon_i$, as required. $\square$

5 A property tester for $([n]^d, \mathcal{D})$

Let $f : [n]^d \to \mathbb{R}$ be the input function, where $[n]^d$ is equipped with a (known) distribution $\mathcal{D} = \mathcal{D}_1 \times \cdots \times \mathcal{D}_d$. The following algorithm is a monotonicity tester for f.

highdim-monotonicity-test $(f, \mathcal{D}, \varepsilon)$

1 `repeat` $O(\varepsilon^{-1} d 2^d)$ `times`
2   `choose random dimension` $j \in [d]$
3   `choose random` $(i_1, \ldots, i_d) \in_{\mathcal{D}} [n]^d$
4   `run one iteration of` *`repeat`* `loop of` **monotonicity-test**$(f^j_{i_1,\ldots,\hat{i_j},\ldots,i_d}, \mathcal{D}_j, *)$
  `output ACCEPT`

Clearly, for $\mathcal{D} = \mathcal{U}$ **highdim-monotonicity-test** is equivalent to **highdim-mon-uniform-test**.

We start with the query complexity analysis. The call to **monotonicity-test** in line 4 has query complexity $O(H_j)$ (the entropy of $\mathcal{D}_j$). Therefore, the expected query complexity in each iteration of the *repeat* loop is $\frac{1}{d}\sum_{j=1}^d O(H_j) = \frac{1}{d}O(H)$ (we use the well known identity that the entropy of a product of independent variables is the sum of the individual entropies). Therefore the total running time is $O(\varepsilon^{-1} 2^d H)$, as claimed.

We prove correctness. Clearly, if f is monotone then **highdim-monotonicity-test** outputs "ACCEPT" with probability 1. Assume f is ε-far from monotone. In order to lower bound the success probability (outputting "REJECT") of line 4, we want to lower bound the average axis-parallel distances to monotonicity of f, similarly to Lemma 4. In order to do that, we consider the continuous case. Recall the definition of the function $\tilde{f} : I^d \to \mathbb{R}$ from Section 2. Let $\tilde{\varepsilon}$ be its distance from monotonicity w.r.t. the Lebesgue measure, and $\tilde{\varepsilon}_i$ its corresponding axis-parallel distances. We need the following lemma, which is a continuous version of Lemma 4.

Lemma 6.

$$\sum_{i=1}^{d} \tilde{\varepsilon}_i \geq \tilde{\varepsilon}/2^{d+1}.$$

Proof. The proof is basically as that of Lemma 4, with a redefinition of $B_i, B, \chi_B, \Psi_L, \Psi_R, \Psi_L^i, \Psi_R^i, \varphi_L^{(i)}, \varphi_R^{(i)}$. We pick an arbitrarily small $\delta > 0$, and define the set $B_i \subseteq I^d$ as the set $\{f \neq g\}$ for some $\leq_i$-monotone g with distance at most $\tilde{\varepsilon}_i + \delta$ from f (so $\tilde{\varepsilon}_i \leq \mu(B_i) \leq \tilde{\varepsilon}_i + \delta$). Let χ_B be the characteristic function of $B = \cup B_i$. Obviously, $\mu(B) \leq \sum \tilde{\varepsilon}_i + \delta d$. We then define the following continuous versions of Ψ_L, Ψ_R, which are now operators on measurable $\{0,1\}$ functions over I:

$$(\Psi_L v)(x) = \begin{cases} 1 & v(x) = 1 \text{ or there exists } y \in [0, x) \text{ s.t. } \int_y^x v(t)dt \geq \frac{1}{2}(x - y) \\ 0 & \text{otherwise} \end{cases}$$

$$(\Psi_R v)(x) = \begin{cases} 1 & v(x) = 1 \text{ or there exists } y \in (x, 1] \text{ s.t. } \int_x^y v(t)dt \geq \frac{1}{2}(y - x) \\ 0 & \text{otherwise} \end{cases}$$

The operator Ψ_L^i (resp. Ψ_R^i) on functions of I^d applies Ψ_L (resp. Ψ_R) on all lines of the form

$$\{x_1\} \times \cdots \times \{x_{i-1}\} \times I \times \{x_{i+1}\} \times \cdots \times \{x_d\}\ .$$

The functions $\varphi_L^{(i)}$ and $\varphi_R^{(i)}$ are defined as in (1). The main observation is that $\mu(\{\varphi_L^{(1)} = 1\}) \leq 2^d \mu(B)$ (similarly, for $\varphi_R^{(1)}$). This is a simple consequence of the following lemma, which is a continuous version of Lemma 5.

Lemma 7. *Let v be a measurable $\{0,1\}$ function defined on I. Then $\int_0^1 (\Psi_L v)(t)dt \leq 2\int_0^1 v(t)dt$. In other words, the measure of the 1-level set of $\Psi_L v$ is at most twice the measure of the 1-level set of v. A similar result holds for Ψ_R.*

The mostly technical proof of Lemma 7 can be found in Appendix B. The rest of the proof of Lemma 6 continues very similar to that of Lemma 4 and by taking $\delta \to 0$.

□

As a result of Lemmas 6 and 1, we have

$$\sum \varepsilon_i \geq \varepsilon/2^{d+1}$$

.

This means that the expected one-dimensional distance from monotonicity of $f^j_{i_1,\ldots,\hat{i_j},\ldots,i_d}$ in line 4 (w.r.t. the marginal distribution $\mathcal{D}_j$) is at least $\gamma = \frac{1}{d}\varepsilon/2^{d+1}$. By the analysis of **monotonicity-test**, we know that the probability of outputting "REJECT" in a single iteration of the repeat loop is $\Omega(\gamma)$. Therefore, by repeating $O(1/\gamma)$ times we get constant probability of success. This completes the proof of Theorem 1. □

6 The Distribution-(almost) Free Case

We prove Theorem 2. Let $f : [n]^d \to \mathbb{R}$ be the input function, where $[n]^d$ is equipped with a distribution $\mathcal{D} = \mathcal{D}_i \times \cdots \times \mathcal{D}_d$, and the marginal distributions $\mathcal{D}_i$ are unknown.

We cannot simply run **highdim-monotonicity-test** on f, because that algorithm expects the argument $\mathcal{D}$ to be the actual probabilities of the distribution. In the distribution-free case, we can only pass an oracle$[\mathcal{D}]$, which is a distribution sampling function. Therefore our new algorithm, **highdim-monotonicity-test-distfree** will take f, oracle$[\mathcal{D}]$ and ε as input.

highdim-monotonicity-test1 $(f, \texttt{oracle}[\mathcal{D}], \varepsilon)$

1 `repeat` $O(\varepsilon^{-1}d2^d)$ `times`
2 `choose random dimension` $j \in [d]$
3 `choose random` $(i_1, \ldots, \hat{i}_j, \ldots, i_d) \in_{\mathcal{D}} [n]^{d-1}$
4 `run one iteration of` *`repeat`* `loop of` **monotonicity-test1**$(f^j_{i_1,\ldots,\hat{i_j},\ldots,i_d}, \texttt{oracle}[\mathcal{D}_j], *)$
`output ACCEPT`

Note that oracle$[\mathcal{D}_j]$ in line 4 is obtained by projecting the output of oracle$[\mathcal{D}]$. Algorithm **monotonicity-test1** is defined to be exactly Halevy-Kushilevitz's 1-dimensional distribution-free monotonicity tester [10]. It is similar to **monotonicity-test**, except that we cannot compute $p_{\text{int}(x)}$ and therefore we cannot compute y and $\text{int}(y)$. Instead, $\text{int}(y)$ is chosen uniformly at random in the integer interval

$$[\text{int}(x), \min\{\text{int}(x) + 2^t\}] ,$$

for $t = 0, \ldots, \lceil \log n \rceil$. This replaces the while-loop starting on line 4, and similarly for the while-loop starting on line 8. The running time of a single iteration of the repeat loop of **monotonicity-test1** is $O(\log n)$, and the total running time is $O(\frac{d2^d \log n}{\varepsilon})$, as required.

Let f' denote the one dimensional function $f^j_{i_1,\ldots,\hat{i_j},\ldots,i_d}$, as chosen in line 4 of **highdim-monotonicity-test1**, and let ε' be its distance from monotonicity w.r.t. $\mathcal{D}_j$. In [10] it is proven that a single repeat-loop iteration of **monotonicity-test1** $(f, \text{oracle}[\mathcal{D}_j], *)$ outputs "REJECT" with probability $\Omega(\varepsilon')$. But we showed in Section 5 that $\mathbf{E}[\varepsilon'] \geq \frac{1}{d}\varepsilon/2^{d+1}$. Repeating lines 2-4 $O(\varepsilon^{-1}d2^d)$ times amplifies this to a constant probability. This concludes the proof of Theorem 2.

□

7 Future work

1. *Lower bounds:* The best known lower bound for the one-dimensional uniform distribution property tester [3] is $\Omega(\varepsilon^{-1}\log n)$. For arbitrary distribution it is possible, using Yao's minimax principal, to show a lower bound of $\Omega(\varepsilon^{-1}\log(\varepsilon/p_{max}))$, where p_{max} is the maximal probability in the distribution. Note that $\log(1/p_{max})$ can be arbitrarily smaller than H. It would be interesting to close the gap, as well as generalize for higher dimension.
2. *High-dimensional monotonicity:* It is not known if Lemma 4 is tight. Namely, is there a high dimensional function that has axis-parallel distances from monotonicity exponentially (in d) smaller than the global distance to monotonicity? We note that even if the exponential dependence is tight in the inequality, it would not necessarily mean that the property testing query complexity should be exponential in d (other algorithms that are not based on axis-parallel comparisons might do a better job).
3. *Other posets and distributions:* It would be interesting to generalize the results here to functions over general posets [6] as well as arbitrary distributions (not necessarily product distributions).
4. *More information theory in property testing:* It would be interesting to see how the entropy or other complexity measures of $\mathcal{D}$ affect the query complexity of other interesting property testing problems.

Acknowledgements

We would like to thank Shirley Halevy and Eyal Kushilevitz for enlightening discussions.

References

1. Batu, T., Rubinfeld, R., White, P. *Fast approximate PCPs for multidimensional bin-packing problems*, Proc. RANDOM (1999), 245–256.
2. Dodis, Y., Goldreich, O., Lehman, E., Raskhodnikova, S., Ron, D., Samorodnitsky, A. *Improved testing algorithms for monotonicity*, Proc. RANDOM (1999), 97–108.
3. Ergun, F., Kannan, S., Kumar, S. Ravi, Rubinfeld, R., Viswanathan, M. *Spot-checkers*, Proc. STOC (1998), 259–268.
4. Fischer, E. *The art of uninformed decisions: A primer to property testing*, Bulletin of EATCS, 75: 97-126, 2001.
5. Fischer, E., Kindler, G., Ron, D., Safra, S., Samorodnitsky, A., *Testing Juntas* Proc. FOCS (2002), 103–112.
6. Fischer, E., Lehman, E., Newman, I., Raskhodnikova, S., Rubinfeld, R., Samorodnitsky, A. *Monotonicity testing over general poset domains*, Proc. STOC (2002), 474–483.
7. Goldreich, O. *Combinatorial property testing - A survey*, in "Randomization Methods in Algorithm Design," 45-60, 1998.
8. Goldreich, O., Goldwasser, S., Lehman, E., Ron, D., Samordinsky, A. *Testing monotonicity*, Combinatorica, 20 (2000), 301–337.
9. Goldreich, O., Goldwasser, S., Ron, D. *Property testing and its connection to learning and approximation*, J. ACM 45 (1998), 653–750.
10. Halevy, S., Kushilevitz, E. *Distribution-free property testing*, Proc. RANDOM (2003), 302–317.
11. Halevy, S., Kushilevitz, E. *Testing Monotonicity over Graph Products*, ICALP (2004).
12. Parnas, M., Ron, D., Rubinfeld, R. *Tolerant property testing and distance approximation*, ECCC 2004.
13. Ron, D. *Property testing*, in "Handbook on Randomization," Volume II, 597-649, 2001.
14. Rubinfeld, R., Sudan, M. *Robust characterization of polynomials with applications to program testing*, SIAM J. Comput. 25 (1996), 647–668.

Appendix A: Proof of Lemma 1

The direction $\tilde{\varepsilon} \le \varepsilon$ is clear. It remains to show that $\varepsilon \le \tilde{\varepsilon}$. Pick an arbitrarily small $\delta > 0$, and let $\tilde{g}$ be some monotone function on I^d with distance at most $\tilde{\varepsilon} + \delta$ to $\tilde{f}$. We are going to replace $\tilde{g}$ with a monotone function g over $[n]^d$ with distance at most $\tilde{\varepsilon} + 2\delta$ to f. To do this, we will make it constant on tiles of the form $\Delta^1_{i_1} \times \Delta^2_{i_2} \times \cdots \times \Delta^d_{i_d}$, paying a price of at most one extra δ. We will do this one dimension at a time.

We show how to do this for the first dimension, and the rest is done similarly. Our goal is to replace $\tilde{g}$ with a monotone function $\tilde{g}^{(1)}$ that has distance at most $\tilde{\varepsilon} + \delta(1 + 1/d)$ from $\tilde{f}$, with the property that it is constant on any line segment of the form $\Delta^1_i \times \{x_2\} \times \cdots \times \{x_d\}$, for any $i \in [n]$ and $x_2, \ldots, x_d \in I$. For every $i \in [n]$, do the following: For every $x_1 \in \Delta^1_i$, consider the restriction of the function $\tilde{g}$ to the $d-1$ dimensional cube $\{x_1\} \times I^{d-1}$. Denote this function by $\tilde{g}_{x_1}(x_2, \ldots, x_d)$. Let $\tilde{\varepsilon}_{x_1}$ denote the distance between $\tilde{g}_{x_1}$ and $\tilde{f}_{x_1}$ (where $\tilde{f}_{x_1}$ is defined similarly to $\tilde{g}_{x_1}$). Let $\gamma = \inf_{x_1 \in \Delta^1_i} \tilde{\varepsilon}_{x_1}$. Pick x_1 such that $\tilde{\varepsilon}_{x_1}$ is at most $\gamma + \delta/d$. We now "smear" the value of $\tilde{g}$ at $(x_1, x_2, \ldots, x_d)$ to $\Delta^1_i \times \{x_2\} \times \cdots \times \{x_d\}$, for all $x_2, \ldots, x_d$. Doing this for all $i = 1, \ldots, n$ produces the function $\tilde{g}^{(1)}$. It is not hard to see that the distance between $\tilde{g}^{(1)}$ and f is at most $\tilde{\varepsilon} + \delta(1 + 1/d)$, and the function $\tilde{g}^{(1)}$ is monotone.

After obtaining $\tilde{g}^{(j)}$, we obtain $\tilde{g}^{(j+1)}$ by repeating the above process for the $(j+1)$-th dimension. It is easy to verify that for $j < d$,

1. If $\tilde{g}^{(j)}$ is monotone then so is $\tilde{g}^{(j+1)}$.
2. If $\tilde{g}^{(j)}$ is constant on $\Delta^1_{i_1} \times \Delta^2_{i_2} \times \cdots \times \Delta^j_{i_j} \times \{x_{j+1}\} \times \cdots \times \{x_d\}$ for all $i_1, \ldots, i_j$ and $x_{j+1}, \ldots, x_d$, then $\tilde{g}^{(j+1)}$ is constant on $\Delta^1_{i_1} \times \Delta^2_{i_2} \times \cdots \times \Delta^{j+1}_{i_{j+1}} \times \{x_{j+2}\} \times \cdots \times \{x_d\}$ for all $i_1, \ldots, i_{j+1}$ and $x_{j+2}, \ldots, x_d$.
3. If the distance between $\tilde{g}^{(j)}$ and $\tilde{f}$ is at most $\tilde{\varepsilon} + j\delta/d$, then the distance between $\tilde{g}^{(j+1)}$ and $\tilde{f}$ is at most $\tilde{\varepsilon} + (j+1)\delta/d$.

Therefore, $\tilde{g}^{(d)}$ is monotone, and it is defined over $[n]^d$ (because it is constant over $\Delta^1_{i_1} \times \cdots \times \Delta^d_{i_d}$). Denote the equivalent function over $([n]^d, \mathcal{D})$ by g. The monotone function g has distance at most $\tilde{\varepsilon} + 2\delta$ from f. The set of possible distances between functions over $([n]^d, \mathcal{D})$ is finite, therefore by choosing δ small enough we obtain a function g which has distance exactly $\tilde{\varepsilon}$ from f. This concludes the proof.

□

Appendix B: Proof of Lemma 7

Let B denote the set $\{x | v(x) = 1\}$, and C denote $\{x | (\Psi_L v)(x) = 1\}$. We want to show that $\mu(C) \le 2\mu(B)$. It suffices to show that for any $\varepsilon > 0$, $\mu(C) \le (2 + \varepsilon)\mu(B)$.

For $y < x$, define

$$\rho(y, x) = \frac{\int_y^x v(t)dt}{y - x} = \frac{\mu(B \cap [y, x])}{\mu([y, x])}.$$

That is, $\rho(y, x)$ is the measure of the set $\{v = 1\}$ conditioned on $[y, x]$.

Pick an arbitrary small $\varepsilon > 0$. Let C_ε be the set of points $x \in I$ such that there exists $y < x$ with $\rho(y, x) > 1/2 - \varepsilon$. For $x \in C_\varepsilon$, we say that y is an *ε-witness* for x if $\rho(y, x) > 1/2 - \varepsilon$. We say that y is a *strong ε-witness* for x if for all $y < z \le x$, $\rho(y, z) > 1/2 - \varepsilon$.

We claim that if $x \in C_\varepsilon$, then there exists a strong ε-witness y for x. Assume otherwise. Let y be any ε-witness for x. Since y is not a strong ε-witness for x, there exists $z : y < z < x$ such that $\rho(y, z) \le 1/2 - \varepsilon$. Let z_0 be the supremum of all such z. Clearly, $y < z_0 < x$ (z_0 cannot be x because then by continuity of ρ we would get $\rho(y, x) \le 1/2 - \varepsilon$). We claim that z_0 is a strong witness for x. Indeed, if for some $z' : z_0 < z' < x$ we had $\rho(z_0, z') \le 1/2 - \varepsilon$, then it would imply $\rho(y, z') \le 1/2 - \varepsilon$, contradicting our choice of the supremum.

For all $x \in C_\varepsilon$, let $y(x)$ be the infimum among all strong ε-witnesses of x. We claim that for $x \ne x'$, the intervals $[y(x), x)$ and $[y(x'), x')$ are either disjoint, or $y(x) = y(x')$. Otherwise, we would have, without loss of generality, $y(x) < y(x')$ with both $x, x' > y(x')$. But then any strong ε-witness for x that is strictly between $y(x)$ and $y(x')$ (which exists) is a strong ε-witness for x', contradicting the choice of $y(x')$.

Therefore, the set $Y = y(C_\varepsilon)$ (the image of $y(\cdot)$) is countable, and for any $y_0 \in Y$ there exists an $x(y_0) > y_0$ which is the supremum over all $x : x > y_0$ such that $y(x) = y_0$. For two distinct $y_1, y_2 \in Y$, the intervals $[y_1, x(y_1))$ and $[y_2, x(y_2))$ are disjoint. Let

$$D = \cup_{y \in Y} [y, x(y)) \ .$$

Clearly, by continuity of ρ, for all $y \in Y$

$$\mu([y, x(y))) \leq \frac{\mu([y, x(y)) \cap B)}{1/2 - \varepsilon} \ .$$

Therefore

$$\mu(D) \leq \frac{\mu(D \cap B)}{1/2 - \varepsilon} \ .$$

We also have that $\mu(\bar{D}) = \mu(D)$ (where $\bar{D}$ is the closure of D), because D is a union of countably many intervals. Therefore,

$$\mu(\bar{D}) \leq \frac{\mu(\bar{D} \cap B)}{1/2 - \varepsilon} \ .$$

By our previous claim $C_\varepsilon \subseteq \bar{D}$, therefore

$$\mu(C_\varepsilon) \leq \frac{\mu(\bar{D} \cap B)}{1/2 - \varepsilon} \ ,$$

and thus

$$\mu(C_\varepsilon \cup (B \backslash \bar{D})) \leq \frac{\mu(B)}{1/2 - \varepsilon} \ .$$

We claim that up to a set of measure zero, C is contained in $C_\varepsilon \cup (B \backslash \bar{D})$. Indeed, if $x \in C$, then $(\Psi_L v)(x) = 1$. Therefore, either there exists $y < x$ such that $\rho(y, x) \geq 1/2$, in which case $x \in C_\varepsilon$, or there does not exist such a y: In this case, by definition of Ψ_L, $x \in B$, and also such an x cannot be in the interior of $\bar{D}$. Therefore, we have $x \in B \backslash \bar{D}$ (unless $x \in \partial D$). But since $\mu(\partial D) = 0$, our claim is proven. We conclude that

$$\mu(C) \leq \frac{\mu(B)}{1/2 - \varepsilon},$$

as desired. $\square$