Electronic Colloquium on Computational Complexity, Report No. 27 (2010)

Testing monotonicity of distributions over general partial orders

Arnab Bhattacharyya*
MIT
abhatt@mit.edu

Eldar Fischer†
Technion
eldar@cs.technion.ac.il

Ronitt Rubinfeld‡
MIT, Tel Aviv University
ronitt@csail.mit.edu

Paul Valiant§
U.C. Berkeley
pvaliant@eecs.berkeley.edu

Abstract

We investigate the number of samples required for testing the monotonicity of a distribution with respect to an arbitrary underlying partially ordered set. Our first result is a nearly linear lower bound for the sample complexity of testing monotonicity with respect to the poset consisting of a directed perfect matching. This is the first nearly linear lower bound known for a natural non-symmetric property of distributions. Testing monotonicity with respect to the matching reduces to testing monotonicity with respect to various other natural posets, showing corresponding lower bounds for these posets also. Next, we show that whenever a poset has a linear-sized matching in the transitive closure of its Hasse digraph, testing monotonicity with respect to it requires $\Omega(\sqrt{n})$ samples. Previous such lower bounds applied only to the total order. We also give upper bounds to the sample complexity in terms of the chain decomposition of the poset. Our results simplify the known tester for the two dimensional grid and give the first sublinear bounds for higher dimensional grids and the Boolean cube.

*Supported by a DOE Computational Science Graduate Fellowship and NSF grants 0514771, 0728645 and 0732334.
†Research supported by an ERC-2007-StG grant number 202405 and by an ISF grant number 1011/06
‡Supported by NSF grants 0732334 and 0728645, Marie Curie Reintegration grant PIRG03-GA-2008-231077, and the ISF grant nos. 1147/09 and 1675/09.
§Supported in part by NSF grant 6916481.

ISSN 1433-8092

1 Introduction

We study the complexity of testing the monotonicity of a distribution over the elements of a partially ordered set. Given a poset $P = (V, \preceq)$, a probability distribution p on the elements of V is said to be *monotone with respect to* P if $x \preceq y$ implies $\mathsf{p}(x) \leq \mathsf{p}(y)$. Monotonicity is a natural property of functions on posets and has been extensively studied in the context of property testing (see [EKK$^+$00, GGL$^+$00, DGL$^+$99, BRW99, AKNS99, FN01, Fis01]). Here, we examine the testability of the monotonicity of distributions, where access to the distribution is given only via samples independently generated according to the distribution. We would like to construct efficient algorithms that take as input a poset P, samples of a distribution p and a parameter $\epsilon \in (0, 1)$, and determine correctly with high probability whether p is monotone with respect to P or is ϵ-far away in L_1 distance from any such monotone distribution.

Monotonicity as a property of probability distributions is interesting for several reasons. First, many naturally arising distributions are monotone or hold motivation for monotonicity testing. For example, it may be hypothesized that the probability for suffering from back problems is monotone increasing with the patient's height. Second, monotone distributions have proven to be quite useful algorithmically. Devroye [Dev91] used monotone distributions to more efficiently generate random variables. In terms of testability, it is known that the testing of several distribution properties becomes provably easier if the distribution is promised to be monotone; see Section 1.1 for more details. Thus, monotonicity is often a desirable property for a distribution to have, and it would be valuable to have efficient algorithms to determine whether a distribution is monotone or is far from being monotone.

Here, we investigate how the sample complexity of testing monotonicity depends on the structure of the underlying poset. Our results fall into two classes. The first set of results gives nearly tight lower bounds on the sample complexity for some classes of posets, showing for example that there are posets with respect to which testing monotonicity requires a nearly linear number of samples. The second set of results provides efficient algorithms for testing monotonicity with respect to certain classes of posets.

1.1 Previous Work

While classical statistical tests, such as the χ^2-test, seem to require a number of samples at least linear in the domain size, recent work motivated by property testing has shown that there are many natural properties of distributions that can be tested with a sublinear sample complexity. Such properties include testing whether a distribution is uniform, whether a joint distribution is independent, and estimating the entropy [BFR$^+$00, BFF$^+$01, AAK$^+$07, BDKR02, GMV06, BS07].

In [BKR04], the problem of testing whether a distribution is monotone is considered with respect to totally ordered domains. It is shown there that testing uniformity can be reduced to testing monotonicity with respect to the total order. Since testing uniformity is known to require $\Omega(\sqrt{n})$ samples [GR00, BFR$^+$00, Pan08] for domains of size n, [BKR04] thus yields a sample complexity lower bound of $\Omega(\sqrt{n})$ for testing monotonicity over the total order. They also provide an algorithm with sample complexity $\tilde{O}(\sqrt{n})$ for testing monotonicity over the total order on n elements. This algorithm can be roughly viewed as a reduction from monotonicity testing to polylogarithmically many uniformity testing problems. [BKR04] further shows an $\tilde{O}(m^{3/2})$ sample complexity algorithm for testing monotonicity over the $m \times m$ grid with the dominance partial order (the product order), and conjectures that the algorithm can be extended to yield an $\tilde{O}(m^{d-1/2})$ sample complexity algorithm for testing monotonicity over the grid $[m]^d$ with dominance order.

As mentioned previously, monotonicity has also been studied because of its role as a natural condition on distributions that makes other properties significantly easier to test. As an example of this phenomenon, consider the problem of testing uniformity. Testing uniformity requires $\Omega(\sqrt{n})$ samples for arbitrary input distributions on n elements, as mentioned previously. On the other hand, [BKR04] shows that $O(1)$ samples suffice for distributions that are known to be monotone with respect to a total order. [RS09] investigates distributions on the d-dimensional boolean cube with the subset order (note that the domain size here is 2^d) and shows that testing the uniformity of monotone distributions over the cube requires only $\tilde{O}(d)$ samples. Adamaszek, Czumaj and Sohler in [ACS10] have recently extended this result to the continuous $[0,1]^d$ cube with the dominance order. There is *no* test with finite sample complexity for testing uniformity of arbitrary distributions on $[0,1]^d$, but conditioning the input distribution to be monotone permits a tester with $O(n)$ sample complexity. Similar dramatic savings are also known for testing the closeness of two distributions [BKR04, Val08], for testing the independence of a joint distribution [BKR04], and for estimating the entropy [BDKR02].

Monotonicity, as a property of functions defined on posets, has been extensively studied in the context of property testing [EKK$^+$00, BRW99, GGL$^+$00, DGL$^+$99, FLN$^+$02]. The complexity of the testers in this setting is naturally quite different, as the value of the function at any given point in the domain can be queried directly.

1.2 Our Results and Techniques

We address the issue of how sample complexity depends on the structure of the poset with respect to which monotonicity is defined. Intuitively, one would imagine that, as the number of edges in the transitive closure of a poset becomes larger, testing monotonicity with respect to the poset requires fewer samples since there are more comparable elements, making it more likely for a tester to detect violation of monotonicity. Although this intuition is not strictly true[1], our results can be viewed as making the intuition rigorous in several interesting special cases.

Reductions from the matching poset. The matching poset denotes the poset whose Hasse digraph is a perfect matching with directed edges. One of our main contributions in this paper is to show that testing monotonicity with respect to the matching poset on n elements requires $n^{1-o(1)}$ samples. This result serves as a basis for a broad class of nearly linear sample complexity bounds for more general posets. Such posets include the outward directed binary tree and all bounded degree connected bipartite digraphs with all edges oriented towards the same color class. More generally, the $n^{1-o(1)}$ lower bound applies to any poset containing an up-set (also known as a monotone nondecreasing set; see Section 2.2 for the definition) consisting of a linear number of disjoint bounded-degree outward-directed stars.

Our proof of the lower bound for the matching poset uses the methods developed in [Val08] for symmetric[2] properties and adapts them to the analysis of the *non-symmetric* monotonicity property. As far as we know, this is the first nearly linear lower bound for a non-symmetric distribution property. Previous known nearly linear sample complexity lower bounds were for estimating the L_1 norm distance between two distributions [Val08] and for estimating the support size of a distribution [RRSS07, Val08], both of which can be regarded as distance estimation problems. Note that in general, estimating distance to a distribution property can be a much harder task than distinguishing those distributions that have the property from those that are far from having it

[1] The outward-directed star and the matching both have $O(n)$ edges in the transitive closure, but as we show in this paper, the former has sample complexity $O(1)$ while the latter requires $n^{1-o(1)}$ samples.

[2] A distribution property is symmetric if it is preserved under arbitrary relabelings of the distribution domain.

[Val08] – for the case of testing uniformity, the complexity is nearly linear for the estimation problem but only $O(\sqrt{n})$ for the standard property testing problem. On the other hand, our result for monotonicity testing with respect to the matching indicates that even for some natural property testing problems, a nearly linear number of samples is necessary.

Posets containing a large matching. The nearly linear lower bound does not hold if the poset can be partitioned into a small number of long chains. Our next result applies to such posets. We show that if a poset P contains a matching of size $\Omega(n)$ in its transitive closure, then for some constant $\epsilon \in (0, 1)$, an ϵ-monotonicity tester with respect to P requires $\Omega(\sqrt{n})$ samples. This result subsumes the lower bound given in [BKR04] for testing monotonicity with respect to the total order. Our lower bound is obtained by constructing two distributions, one monotone and the other far from being monotone, that are statistically indistinguishable by a tester using $o(\sqrt{n})$ samples.

Sample complexity in terms of chain decomposition. A chain decomposition of a poset is a partitioning of the poset into disjoint chains. We show that if a poset P can be decomposed into w disjoint chains each of length at most ℓ, then for any constant $\epsilon \in (0, 1)$ there exists an ϵ-monotonicity tester for P requiring only $\tilde{O}(w\sqrt{\ell})$ samples. This result implies, for instance, that testing monotonicity with respect to the poset $[m]^d$ (with the dominance order, where d is fixed and m is growing) requires $\tilde{O}(m^{d-1/2})$ samples, settling a conjecture from [BKR04]. In the case of $d = 2$, our results greatly simplify, and improve by polylogarithmic factors, the result of [BKR04]. We also obtain the first sublinear sample upper bounds for testing monotonicity with respect to the hypercube.

1.3 Preliminaries

The notation $P = (V, \preceq)$ denotes the partial order P obtained by ordering the set V according to a reflexive, antisymmetric, and transitive binary relation $\preceq$ over V. Probability distributions p and q are said to be *ϵ-far from each other* if their L_1-distance is at least ϵ, that is, $\|\mathsf{p} - \mathsf{q}\|_1 = \sum_x |\mathsf{p}(x) - \mathsf{q}(x)| \geq \epsilon$. Recall that, given a poset $P = (V, \preceq)$, a distribution p on V is said to be *monotone with respect to P* if for all $x, y \in V$, $x \preceq y$ implies $\mathsf{p}(x) \leq \mathsf{p}(y)$. p is *ϵ-far from being monotone with respect to P* if p is ϵ-far from any distribution q on V that is monotone with respect to P. Formally, our testing problem is defined as follows:

Definition 1 *Given a poset $P = (V, \preceq)$, a positive integer k, and a constant $\epsilon \in (0, 1)$, an algorithm T is said to be an* ϵ-tester for monotonicity with respect to P with sample complexity k *if for any distribution p on V, the algorithm T, given k independent samples taken from p as input, (i) accepts with probability at least $\frac{2}{3}$ if p is monotone with respect to P, and (ii) accepts with probability at most $\frac{1}{3}$ if p is ϵ-far from being monotone with respect to P. The behavior of T is unspecified when p is neither monotone nor ϵ-far from monotone.*

Most of our mathematical notation is standard. Posets are often identified with the digraph given by their Hasse diagram without any comment. In Section 2.1, we use some concepts from [Val08]. For convenience, we reproduce the definition of the moments of a distribution here:

Definition 2 *Given a positive integer k, positive integers a and b, and distributions $\mathsf{p}_1, \mathsf{p}_2$ over a set V, the* k-based (a, b)-moment *of $(\mathsf{p}_1, \mathsf{p}_2)$, denoted $m_{k,\mathsf{p}_1,\mathsf{p}_2}(a, b)$, equals $k^a k^b \sum_{x \in V} \mathsf{p}_1(x)^a \mathsf{p}_2(x)^b$.*

Observe that the k-based (a, b)-moment of $(\mathsf{p}_1, \mathsf{p}_2)$ is the expected outcome of the following "collision statistic" experiment. Get k independent samples $x_1, \ldots, x_k$ from the distribution p_1 and k

independent samples $y_1, \ldots, y_k$ from the distribution p_2, and count the number of pairs of index tuples $i_1, \ldots, i_a \in [k]$ and $j_1, \ldots, j_b \in [k]$ such that $x_{i_1} = \cdots = x_{i_a} = y_{j_1} = \cdots = y_{j_b}$.

2 Lower Bounds

As discussed in Section 1.1, [BKR04] showed that testing monotonicity over the total order requires $\Omega(\sqrt{n})$ samples, which is known to be tight up to polylogarithmic factors. Here, our main result is a nearly linear (and hence, nearly tight) lower bound for testing monotonicity with respect to another natural poset, the matching. The nearly linear lower bound is also extended to other natural posets. We also generalize the $\Omega(\sqrt{n})$ lower bound to a much larger class of posets.

2.1 Testing monotonicity with respect to the matching

We begin by defining the matching poset formally.

Definition 3 *For integer $n \geq 1$, $M_n = (V, \preceq)$, the* matching poset*, is defined as follows. V is a set of $2n$ elements, $\{x_i : i \in [n]\} \cup \{y_i : i \in [n]\}$. The order relation $\preceq$ is given by $x_i \prec y_i$ for every $i \in [n]$; any other two non-identical elements of V are incomparable.*

A simple reduction from testing the identity of two distributions gives a $\Omega(n^{2/3})$ lower bound for the sample complexity of testing monotonicity with respect to M_n. We provide it here as a warm-up to what follows.

Claim 4 *Let p be a probability distribution on the vertices of M_n. To test if p is monotone or ϵ-far from being monotone with respect to M_n requires $\Omega(n^{2/3})$ samples.*

Proof We show that if there is an ϵ-tester for monotonicity over M_n that makes $o(n^{2/3})$ samples, then there is a tester making $o(n^{2/3})$ samples that distinguishes identical distributions from distributions that are $\Theta(\epsilon)$ apart in statistical distance. This contradicts the sample complexity lower bound for the latter problem that was proved in [Val08], thus showing that a tester as above cannot exist.

Suppose that we have an ϵ-tester for monotonicity over M_n using $q = o(n^{2/3})$ samples, and we want to test whether a pair of distributions $(\mathsf{p}_1, \mathsf{p}_2)$ are identical or are 4ϵ-far from each other. Define the distribution p on M_n as follows: for all $i \in [n]$, $\mathsf{p}(x_i) = \frac{1}{2}\mathsf{p}_1(x_i)$ and $\mathsf{p}(y_i) = \frac{1}{2}\mathsf{p}_2(y_i)$. If $\mathsf{p}_1 = \mathsf{p}_2$, then clearly p is monotone on M_n. On the other hand, if there is statistical distance greater than 4ϵ between p_1 and p_2, then $\sum_{i:\mathsf{p}_1(i) \geq \mathsf{p}_2(i)} \mathsf{p}_1(i) - \mathsf{p}_2(i) > 2\epsilon$. Hence, p is ϵ-far from being monotone on M_n. Thus, we can use the ϵ-tester for monotonicity on M_n, where we sample from p by tossing a fair coin and then sampling from p_1 or p_2 accordingly. ■

Next we prove a much stronger, nearly linear, lower bound.

Theorem 5 *Let p be a probability distribution on V. There exists a constant ϵ_0 such that for any $\epsilon \in (0, \epsilon_0)$, every ϵ-tester for monotonicity with respect to M_n requires $n^{1-o(1)}$ samples.*

Proof First, we present a simple structural claim that characterizes the distributions that are ϵ-far from monotone with respect to the matching:

Claim 6 *A distribution* p *on* V *is* ϵ*-far from monotone if and only if*

$$\sum_{(x_i,y_i):\mathsf{p}(x_i)>\mathsf{p}(y_i)} (\mathsf{p}(x_i) - \mathsf{p}(y_i)) > \epsilon$$

Next, for any constant $\alpha \in (0,1)$, define the following property on pairs of distributions:

$$P_\alpha = \{(\mathsf{p}_1,\mathsf{p}_2) : \mathsf{p}_1,\mathsf{p}_2 \text{ are distributions on } [n] \text{ and } \forall i \in [n], \mathsf{p}_2(y_i) \geq \alpha \cdot \mathsf{p}_1(x_i)\}$$

Our overall strategy is the following. First, we show that when $\epsilon < 1/6$, there is a reduction to ϵ-testing monotonicity with respect to M_n from the problem of distinguishing between distribution pairs that satisfy $P_{1/2}$ and distribution pairs that are ϵ'-far[3] from $P_{1/4}$, where ϵ' is only a function of ϵ (notice that $P_\alpha \subset P_\beta$ for $\alpha > \beta$, so that the statement of the reduction makes sense). The reason that the reduction is helpful is that P_α is a symmetric property: relabeling the elements of $[n]$ does not change whether $(\mathsf{p}_1,\mathsf{p}_2)$ is a member of P_α or not. Therefore, the technology developed in [Val08] is potentially applicable to lower bounding the sample complexity of distinguishing between being in $P_{1/2}$ and being ϵ'-far from $P_{1/4}$. But there is still a problematic feature of the P_α property due to which we cannot apply the results of [Val08] directly. Namely, the family of properties P_α is not "continuous" enough. A distribution pair infinitesimally far from a distribution pair in P_α might not, in fact, be in $P_{\alpha'}$ for any $\alpha' > 0$. In general, for distribution property families that are not continuous, it is known that the techniques from [Val08] do not yield tight bounds. However, for the special case of P_α, we show that it is still possible to suitably modify the techniques from [Val08] and get the desired lower bound.

The precise statement of the reduction is given by the following lemma:

Lemma 7 *For any* $\alpha \in (0,1)$*, there is a constant* $c > 1$ *such that for any* $\epsilon < \frac{\alpha}{2(1+\alpha)}$*, if there is an* ϵ*-tester for monotonicity with respect to* M_n *that makes* $q(n,\epsilon)$ *samples, then there is a tester that, given distributions* $(\mathsf{p}_1,\mathsf{p}_2)$*, makes* $q(n,\epsilon/c)$ *samples and distinguishes between the case that* $(\mathsf{p}_1,\mathsf{p}_2) \in P_\alpha$ *and the case that* $(\mathsf{p}_1,\mathsf{p}_2)$ *is* ϵ*-far from being in* P_β*, where* $\beta = (\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2})/(\frac{1}{1+\alpha} + \frac{\epsilon}{2}) < \alpha$.

Proof Given the distribution pair $(\mathsf{p}_1,\mathsf{p}_2)$, we define a map T_α that takes $(\mathsf{p}_1,\mathsf{p}_2)$ to a distribution p on V. Specifically, for every $i \in [n]$, $\mathsf{p}(x_i) = \frac{\alpha}{1+\alpha}\mathsf{p}_1(i)$ and $\mathsf{p}(y_i) = \frac{1}{1+\alpha}\mathsf{p}_2(i)$.

- If $(\mathsf{p}_1,\mathsf{p}_2) \in P_\alpha$, then p is monotone on M_n because for every $i \in [n]$, $\mathsf{p}(y_i) = \frac{1}{1+\alpha}\mathsf{p}_2(i) \geq \frac{\alpha}{1+\alpha}\mathsf{p}_1(i) = \mathsf{p}(x_i)$.
- Assume that p is ϵ-close to being monotone with respect to M_n. By the definition of T_α, $\sum_i \mathsf{p}(x_i) = \frac{\alpha}{1+\alpha}$ and $\sum_i \mathsf{p}(y_i) = \frac{1}{1+\alpha}$. Let p' be a monotone distribution on M_n that is closest in statistical distance to p. We construct our specific p' along the lines of Claim 6: for each edge (x_i,y_i) that was violated by p, we define $\mathsf{p}'(x_i) = \mathsf{p}'(y_i) = \frac{1}{2}(\mathsf{p}(x_i)+\mathsf{p}(y_i))$, and keep the original values of p everywhere else. Therefore, $\sum_i \mathsf{p}'(x_i) \geq \frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}$ and $\sum_i \mathsf{p}'(y_i) \leq \frac{1}{1+\alpha} + \frac{\epsilon}{2}$. Define a new pair of distributions $(\mathsf{p}'_1,\mathsf{p}'_2)$ by $\mathsf{p}'_1(i) = \frac{\mathsf{p}'(x_i)}{\sum_i \mathsf{p}'(x_i)}$ and $\mathsf{p}'_2(i) = \frac{\mathsf{p}'(y_i)}{\sum_i \mathsf{p}'(y_i)}$ for each $i \in [n]$. Note that

$$\frac{\mathsf{p}'_2(i)}{\mathsf{p}'_1(i)} = \frac{\mathsf{p}'(y_i)}{\mathsf{p}'(x_i)}\frac{\sum_i \mathsf{p}'(x_i)}{\sum_i \mathsf{p}'(y_i)} \geq \frac{\sum_i \mathsf{p}'(x_i)}{\sum_i \mathsf{p}'(y_i)} \geq \frac{\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}}{\frac{1}{1+\alpha} + \frac{\epsilon}{2}}$$

[3]We define the distance between two distribution pairs $(\mathsf{p}_1,\mathsf{p}_2)$ and $(\mathsf{q}_1,\mathsf{q}_2)$ as $\|\mathsf{p}_1 - \mathsf{q}_1\|_1 + \|\mathsf{p}_2 - \mathsf{q}_2\|_1$. Farness from P_α is measured using this notion of distance.

so $(\mathsf{p}'_1, \mathsf{p}'_2) \in P_\beta$. Moreover,

$$|\mathsf{p}'_1 - \mathsf{p}_1| = \sum_{i:\mathsf{p}_2(i) \geq \alpha \mathsf{p}_1(i)} \left(\frac{\mathsf{p}(x_i)}{\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}} - \frac{\mathsf{p}(x_i)}{\frac{\alpha}{1+\alpha}} \right) + \sum_{i:\mathsf{p}_2(i) < \alpha \mathsf{p}_1(i)} \left| \frac{\mathsf{p}(x_i)}{\frac{\alpha}{1+\alpha}} - \frac{\mathsf{p}'(x_i)}{\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}} \right|$$
$$\leq \frac{\frac{\epsilon}{2} \sum_i \mathsf{p}(x_i)}{(\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}) \frac{\alpha}{1+\alpha}} + \sum_{i:\mathsf{p}_2(i) < \alpha \mathsf{p}_1(i)} \frac{\mathsf{p}(x_i) - \mathsf{p}'(x_i)}{\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}} \leq 2 \frac{\epsilon/2}{\frac{\alpha}{1+\alpha} - \frac{\epsilon}{2}}$$

Similarly,

$$|\mathsf{p}'_2 - \mathsf{p}_2| = \sum_{i:\mathsf{p}_2(i) \geq \alpha \mathsf{p}_1(i)} \left(\frac{\mathsf{p}(y_i)}{\frac{1}{1+\alpha}} - \frac{\mathsf{p}(y_i)}{\frac{1}{1+\alpha} + \frac{\epsilon}{2}} \right) + \sum_{i:\mathsf{p}_2(i) < \alpha \mathsf{p}_1(i)} \left| \frac{\mathsf{p}(y_i)}{\frac{1}{1+\alpha}} - \frac{\mathsf{p}'(y_i)}{\frac{1}{1+\alpha} + \frac{\epsilon}{2}} \right|$$
$$\leq \frac{\frac{\epsilon}{2} \sum_i \mathsf{p}(y_i)}{(\frac{1}{1+\alpha} + \frac{\epsilon}{2}) \frac{1}{1+\alpha}} + \sum_{i:\mathsf{p}_2(i) < \alpha \mathsf{p}_1(i)} \frac{\mathsf{p}'(y_i) - \mathsf{p}(y_i)}{\frac{1}{1+\alpha}} \leq 2 \frac{\epsilon/2}{\frac{1}{1+\alpha}}$$

Using the condition that $\epsilon < \frac{\alpha}{2(1+\alpha)}$, we have that $|\mathsf{p}_1 - \mathsf{p}'_1| + |\mathsf{p}_2 - \mathsf{p}'_2|$ is $c\epsilon$-close to P_β where c depends only on α.

The lemma is now immediate. To sample from p, one can toss a coin that is biased to be heads with probability $\frac{\alpha}{1+\alpha}$ and then sample from p_1 if the coin comes up heads and from p_2 otherwise. ■

To prove this lemma, we leverage and extend machinery from [Val08], specifically, the following corollary of the Wishful Thinking theorem stated in [Val07].

Theorem 8 (Corollary 1 in [Val07]) *Suppose we are given two distribution pairs $(\mathsf{p}_1, \mathsf{p}_2)$ and $(\mathsf{q}_1, \mathsf{q}_2)$ where the distributions are over $[n]$, a real number $\rho \in \left(0, \frac{1}{10 \cdot 2^{\sqrt{\log n}}}\right)$, and a positive integer k such that the maximum probability assigned by any of the distributions $\mathsf{p}_1, \mathsf{p}_2, \mathsf{q}_1, \mathsf{q}_2$ to a single element is at most $\frac{\rho}{k}$. If*

$$40\rho + 10 \sum_{a,b: 2 \leq a+b \leq \sqrt{\log n}} |m_{k,\mathsf{p}_1,\mathsf{p}_2}(a,b) - m_{k,\mathsf{q}_1,\mathsf{q}_2}(a,b)| < .01$$

then it is impossible to test using k samples any property that is true for $(\mathsf{p}_1, \mathsf{p}_2)$ and false for $(\mathsf{q}_1, \mathsf{q}_2)$.

The observation motivating the above theorem is that for symmetric properties, essentially all that a tester can do to distinguish a distribution pair satisfying the property from a distribution pair not satisfying the property, is to look at the collision statistics of samples from the two distribution pairs. Although this observation was made explicitly in earlier work such as [BDKR02], [Val08] made rigorous the connection between the collision statistics and the values of the moments for distributions with no large weight. Thus, in order to show our lower bound, we need to describe a distribution pair $(\mathsf{p}_1, \mathsf{p}_2) \in P_{1/2}$ and a distribution pair $(\mathsf{q}_1, \mathsf{q}_2)$ that is far from $P_{1/4}$, such that the following two conditions are met: (i) none of the distributions assigns large weight to any element, and (ii) $|m_{k,\mathsf{p}_1,\mathsf{p}_2}(a,b) - m_{k,\mathsf{q}_1,\mathsf{q}_2}(a,b)|$ is small for each (a,b) with $2 \leq a+b \leq \sqrt{\log n}$. Our strategy will be to start from candidate distribution pairs $(\mathsf{p}_1, \mathsf{p}_2) \in P_{1/2}$ and $(\mathsf{q}_1, \mathsf{q}_2)$ that is far from $P_{1/4}$, that satisfy condition (i), and then modify them by a small amount (relative to the L_1 norm) so that condition (ii) is satisfied. Note that the modification of $(\mathsf{p}_1, \mathsf{p}_2)$ needs to be still in $P_{1/2}$ and the modification of $(\mathsf{q}_1, \mathsf{q}_2)$ needs to be still far from $P_{1/4}$. This is where the "delicacy" of monotonicity becomes an issue: even a small change in the L_1 norm to $(\mathsf{p}_1, \mathsf{p}_2)$ could potentially make $\mathsf{p}_2(i) = 0$ for some i. However note that there is already a certain laxity inherent in $P_{1/2}$; namely, we can make $\mathsf{p}_1(i) + \mathsf{p}_2(i)$ as small or as large as we want, as long as $\mathsf{p}_2(i)$ stays at least $\mathsf{p}_1(i)/2$. These issues are at the heart of the following lemma.

Lemma 9 *For any positive integer n, integer $k \in [100^{\sqrt{\log n}}, n-1]$ and weight $w \in (1/n, 1)$, there exist $\{m_{a,b} \ : \ (a,b) \in \{0,1,\ldots,\sqrt{\log n}\}^2\}$, such that given any distribution pair $\mathsf{p} = (\mathsf{p}_1, \mathsf{p}_2)$ on n elements such that $\mathsf{p}_i(j) < \frac{1}{k}$ for all $i \in \{1,2\}$ and $j \in [n]$, there is a distribution pair $\bar{\mathsf{p}} = (\bar{\mathsf{p}}_1, \bar{\mathsf{p}}_2)$ on n elements with the following properties:*

1. *If $\mathsf{p} \in P_{1/2}$ then $\bar{\mathsf{p}} \in P_{1/2}$ also.*
2. $|\mathsf{p}_1 - \bar{\mathsf{p}}_1| + |\mathsf{p}_2 - \bar{\mathsf{p}}_2| \leq w$.
3. *For any $a, b \leq \sqrt{\log n}$, setting $\bar{k} = \frac{kw}{100 \cdot 2^{21\sqrt{\log n}}}$, each $\bar{k}$-based (a,b)-moment $\sum_i \bar{\mathsf{p}}_1^a(i)\bar{\mathsf{p}}_2^b(i)\bar{k}^{a+b}$ is within a $\frac{1}{6000 \log n}$ difference of $m_{a,b}$,*
4. *$\bar{\mathsf{p}}_1(i), \bar{\mathsf{p}}_2(i) \leq (2^{14\sqrt{\log n}} \cdot \bar{k})^{-1}$ for all $i \in [n]$, for $\bar{k}$ as defined above*

Proof The proof of this lemma is similar to the Matching Moments Theorem of [Val08]. The difference is condition (1) above which requires that the transformation from p to $\bar{\mathsf{p}}$ preserves membership in $P_{1/2}$. This condition introduces more technical difficulties. For completeness, we will give all the details. Here are the steps of the transformation from p to $\bar{\mathsf{p}}$, along with explanations of why do they work and why are they well-defined.

1. Set $w' = \frac{w}{7}$. Define I to be the set of $\lfloor w'n \rfloor$ columns[4] i with the smallest value of $\mathsf{p}_1(i) + \mathsf{p}_2(i)$. For each $i \in I$, set $\mathsf{p}_1(i)$ and $\mathsf{p}_2(i)$ to 0. For the remaining columns, modify p such that:

   (i) For each $j \in \{1,2\}$, $\sum_{i \notin I} \mathsf{p}_j(i) = 1 - w'$

   (ii) p changes by at most $3w'$ in L_1 distance

   (iii) If $\mathsf{p} \in P_{1/2}$ before the modification, then after the modification also, for $i \notin I$, $\mathsf{p}_2(i) \geq \frac{1}{2}\mathsf{p}_1(i)$

   (iv) For each $j \in \{1,2\}$ and $i \notin I$, $\mathsf{p}_j(i) < 1/k$.

   We explain how this can be achieved. Assume that $\mathsf{p} \in P_{1/2}$. The columns of p not in I have weight at least $(2-2w')$ by Markov's inequality, so each row has weight at least $1-2w'$ (and at most 1) in the columns not in I. First modify p_1 in the following way. If $\sum_{i \notin I} \mathsf{p}_1(i) < 1 - w'$, then for indices i such that $\mathsf{p}_1(i) < \frac{1}{n}$, increase their weights, maintaining $\mathsf{p}_1(i) \leq \frac{1}{n}$, until $\sum_{i \notin I} \mathsf{p}_1(i) = 1 - w'$. Otherwise, if $\sum_{i \notin I} \mathsf{p}_1(i) > 1 - w'$, decrease weights from arbitrary columns until $\sum_{i \notin I} \mathsf{p}_1(i) = 1 - w'$. In this process, we could have added weight at most w' to this row, and it is still true that $\mathsf{p}_1(i) < \frac{1}{k}$ for all $i \in [n]$. Now, for each $i \notin I$ such that $\mathsf{p}_2(i) < \frac{1}{2}\mathsf{p}_1(i)$, make $\mathsf{p}_2(i) = \mathsf{p}_1(i)/2$; this adds weight at most $w'/2$ to p_2. If $\sum_{i \notin I} \mathsf{p}_2(i) >$

[4] The ith column of p refers to the two-element vector $\langle \mathsf{p}_1(i), \mathsf{p}_2(i) \rangle$.

$1-w'$, remove weight from the columns i such that $\mathsf{p}_2(i) > \mathsf{p}_1(i)$ subject to the restriction $\mathsf{p}_2(i) \geq \mathsf{p}_1(i)$, until $\sum_{i \notin I} \mathsf{p}_2(i) = 1-w'$. Otherwise, if $\sum_{i \notin I} \mathsf{p}_2(i) < 1-w'$, for indices i such that $\mathsf{p}_2(i) < \frac{1}{n}$, increase their weights, maintaining $\mathsf{p}_2(i) \leq \frac{1}{n}$, until $\sum_{i \notin I} \mathsf{p}_2(i) = 1-w'$. p_1 moves by at most w' while p_2 moves by at most $2w'$ in the L_1 distance.

This stage ensures that now there are $\lfloor w'n \rfloor$ columns of zeros, corresponding to I, and outside of them both the entries of p_1 and the entries of p_2 sum up to $1-w'$.

2. Let $\mu = 1 + \lfloor \sqrt{\log n} \rfloor$ and $\lambda = \frac{kw'}{6\mu^5 \cdot 60^\mu}$. For integers $0 \leq a, b \leq \mu - 1$ such that $a + b \geq 2$, let $\sigma_{a,b} = \sum_{i \notin I} \mathsf{p}_1^a(i) \mathsf{p}_2^b(i) \lambda^{a+b}$; additionally, set $\sigma_{a,b} = 0$ for $a + b < 2$.

   For an interval in the integers $[u, v]$, let $\ell^{[u,v]}$ be defined as the matrix with entries j^i for columns indexed by $j \in [u, v]$ and rows indexed by $i \in [0, v-u]$. For an integer $\mu > 1$, let $L^\mu = \ell^{[1,\mu]} \otimes \ell^{[\mu+1,2\mu]}$, where $\otimes$ denotes the tensor product operation[5] Define c to be $(L^\mu)^{-1}\sigma$ (note that here we refer to σ as a vector whose coordinates are indexed by number pairs, and multiply it by a matrix whose columns are indexed by pairs). The coordinates of this vector are indexed by pairs (γ, δ) where $\gamma \in [1, \mu]$ and $\delta \in [\mu+1, 2\mu]$.

3. Let $\bar{\sigma}_{a,b}$ have value 0 when $a + b < 2$ and value $\frac{\lambda^2}{k}$ otherwise. Note that $\bar{\sigma}_{a,b}$ is an upper bound on $\sigma_{a,b}$ for each value of a, b, because each entry of p is bounded by $\frac{1}{k}$, and so $\sigma_{a,b}$ is maximized when k columns of p equal $\frac{1}{k}\langle 1, 1 \rangle$. Let $\bar{L}^\mu$ be an element-by-element upper bound on the magnitudes of the elements in $(L^\mu)^{-1}$; by the claim below, all its entries can be 60^μ. Let $\bar{c} = \bar{L}^\mu \cdot \bar{\sigma}$ be a vector that upper bounds each entry of c. Each entry of $\bar{c}$ equals $(\mu^2 - 3) \cdot 60^\mu \cdot \frac{\lambda^2}{k}$.

   Claim 10 *Each element of $(L^\mu)^{-1}$ is at most 60^μ in absolute value.*

   Proof Matrix inversion and tensor product commute: $(L^\mu)^{-1} = (\ell^{[1,\mu]})^{-1} \otimes (\ell^{[\mu+1,2\mu]})^{-1}$. $\ell^{[1,\mu]}$ and $\ell^{[\mu+1,2\mu]}$ are Vandermonde matrices and the entries of their inverses can be bounded by a formula from [Kli67], cited in [Val08]. Using this, we find that each entry of $(\ell^{[1,\mu]})^{-1} \otimes (\ell^{[\mu+1,2\mu]})^{-1}$ is at most $(2e)^\mu \cdot (4e)^\mu \leq 60^\mu$ in magnitude. ■

4. For each $\gamma \in [1, \mu]$ and $\delta \in [\mu+1, 2\mu]$, choose $\lfloor \bar{c}_{\gamma,\delta} - c_{\gamma,\delta} \rfloor$ many of the zeroed-out columns of p (those supported by I) and make all of them $\langle \frac{\gamma}{\lambda}, \frac{\delta}{\lambda} \rangle$. Note that $\mathsf{p}_2(i) > \mathsf{p}_1(i)$ for any such column i, and hence membership in $P_{1/2}$ is not affected.

   For this modification to make sense, we need to make sure that the total number of columns changed is less than $|I| = \lfloor w'n \rfloor$ and the total weight added to each of the two rows is less than w'. The total weight added to the p_1 row is $\sum_{\gamma,\delta} \lfloor \bar{c}_{\gamma,\delta} - c_{\gamma,\delta} \rfloor \frac{\gamma}{\lambda}$. Note that $\sum_{\gamma,\delta} c_{\gamma,\delta}\gamma = (L^\mu c)_{1,0} = \sigma_{1,0} = 0$. Therefore:

$$\sum_{\gamma,\delta} \lfloor \bar{c}_{\gamma,\delta} - c_{\gamma,\delta} \rfloor \frac{\gamma}{\lambda} \leq \sum_{\gamma,\delta} \bar{c}_{\gamma,\delta} \frac{\gamma}{\lambda} \leq (\mu^2 - 3) 60^\mu \frac{\lambda^2}{k} \sum_{\gamma,\delta} \frac{\gamma}{\lambda} \leq \frac{(\mu^2-3)w'}{6\mu^5} \frac{\mu^2(\mu+1)}{2} \leq \frac{(\mu^2-3)w'}{6\mu^2} \leq \frac{w'}{6}.$$

   Similarly for the p_2 row, $\sum_{\gamma,\delta} \lfloor \bar{c}_{\gamma,\delta} - c_{\gamma,\delta} \rfloor \frac{\delta}{\lambda} \leq \frac{(\mu^2-3)w'}{4\mu^2} \leq \frac{w'}{4}$. The total number of columns changed is at most $\sum_{\gamma,\delta} (\bar{c}_{\gamma,\delta} - c_{\gamma,\delta})$. Observe that $\sum_{\gamma,\delta} c_{\gamma,\delta} = (L^\mu c)_{0,0} = \sigma_{0,0} = 0$. Then, $\sum_{\gamma,\delta} (\bar{c}_{\gamma,\delta} - c_{\gamma,\delta}) = \sum_{\gamma,\delta} \bar{c}_{\gamma,\delta} = (\mu^2 - 3) 60^\mu \frac{\lambda^2}{k} \mu^2 = \frac{(\mu^2-3)w'\lambda}{6\mu^3} = w' \frac{kw'}{36\mu^6 60^\mu} \leq w' \frac{n}{36} \leq \lfloor w'n \rfloor$.

[5] Given a real matrix X with rows and columns indexed respectively by i and j, and a real matrix Y indexed by k and l, the tensor product $X \otimes Y$ is defined to be the matrix with rows indexed by pairs (i, k), columns indexed by pairs (j, l), and the entry at $((i, k), (j, l))$ given by $X(i, j) \cdot Y(k, l)$.

5. Make $\sum_i \mathsf{p}_1(i) = 1$ by filling in the columns that were not assigned in the previous step with equal weights. Do the same for p_2.

   We show that for any column i filled in during this step, $\mathsf{p}_2(i) > \frac{1}{2}\mathsf{p}_1(i)$. Let $x = \frac{\mu^2-3}{\mu^2}w'$. Redoing the calculations in the above part to take account of the floors, we find that the weight added to the p_1 row in step 4 is in the interval $[\frac{x}{6} - \frac{\mu^3}{\lambda}, \frac{x}{6}]$ and that the weight added to the p_2 row is in the interval $[\frac{x}{4} - \frac{2\mu^3}{\lambda}, \frac{x}{4}]$. So, the weight that the current step adds to p_1 is in the interval $[(\frac{5}{6}+\frac{3}{6\mu^2})w', (\frac{5}{6}+\frac{3}{6\mu^2})w' + \frac{\mu^3}{\lambda}]$, while that added to p_2 is in $[(\frac{3}{4}+\frac{3}{4\mu^2})w', (\frac{3}{4}+\frac{3}{4\mu^2})w' + \frac{2\mu^3}{\lambda}]$. Also, the number of columns filled in step 4 is in the interval $[\frac{x\lambda}{6\mu} - \mu^2, \frac{x\lambda}{6\mu}]$, and the current step fills the rest of the $\lfloor w'n \rfloor$ columns. So, the minimum ratio between $\mathsf{p}_2(i)$ and $\mathsf{p}_1(i)$ for a column i filled during this step is at least:

$$\frac{(\frac{3}{4}+\frac{3}{\mu^2})w'}{(\frac{5}{6}+\frac{3}{\mu^2})w' + \frac{\mu^3}{\lambda}} \geq \frac{9}{10} - o(1) > \frac{1}{2}$$

We define $\bar{\mathsf{p}}$ to be the distribution pair that results after these five modification steps to p. It remains to show that all four claims made in the lemma hold. If originally $\mathsf{p} \in P_{1/2}$, then $\bar{\mathsf{p}} \in P_{1/2}$ also for the reasons explained above. To bound the distance that p moves during the modifications, note that p changes by at most $5w'$ during step 1 ($2w'$ when the columns in I are zeroed-out and $3w'$ for making the rest of the columns in each row add up to $1-w'$) and by $2w'$ in the rest of the steps (since that much weight is added to the zeroed-out columns), making the total distance moved at most $7w' = w$.

Now, as for the $\bar{k}$-based moments of $\bar{p}$, observe that the $(0,0)$ moment is exactly n and the $(0,1)$ and $(1,0)$ moments are exactly $\bar{k}$. There is however some variation in the $\bar{k}$-based (a,b) moments for $a+b \geq 2$. For ease of analysis, we first bound the variation in the λ-based moments and then scale to $\bar{k}$-based moments.

We define now $L^\mu_{a,b}$ to be the (a,b)-row of L^μ, and set $m_{a,b} = L^\mu_{a,b} \cdot \bar{c}$. [6] Observe that the (a,b)-moment contributed by the indices not in I is exactly $\sigma_{a,b}$, while the (a,b)-moment contributed by the columns set in the fourth step is $L^\mu_{a,b} \cdot \lfloor \bar{c} - c \rfloor \in [L^\mu_{a,b} \cdot (\bar{c}-1) - \sigma_{a,b}, L^\mu_{a,b} \cdot \bar{c} - \sigma_{a,b}]$.[7] To analyze the moments contributed by the weights added in the fifth step, recall that in this step, the number of entries allocated is $[\lfloor w'n \rfloor - \frac{x\lambda}{6\mu}, \lfloor w'n \rfloor - \frac{x\lambda}{6\mu} + \mu^2]$ and the weight added is $[(\frac{3}{4}+\frac{3}{4\mu^2})w', (\frac{3}{4}+\frac{3}{4\mu^2})w' + \frac{2\mu^3}{\lambda}]$ for p_2 and $[(\frac{5}{6}+\frac{3}{6\mu^2})w', (\frac{5}{6}+\frac{3}{6\mu^2})w' + \frac{\mu^3}{\lambda}]$ for p_1. The ratio between the minimum and maximum contributions of step 5 to the (a,b) λ-based moment can then be lower-bounded by $1 - (a+b)\frac{6\mu^3}{w'\lambda}$.[8] On the other hand, the maximum (a,b) λ-based moment contribution can be upper-bounded by $\frac{(0.9w')^{a+b}}{(\frac{w'n}{2})^{a+b-1}}\lambda^{a+b} \leq 2^{a+b}w'\lambda\left(\frac{\lambda}{n}\right)^{a+b-1}$. Then, for $a+b \in [2,\mu]$, the difference between the maximum and minimum contributions to the (a,b) λ-based moments contributed by the weights added in step 5 can be upper-bounded by $(a+b)\frac{6\mu^3}{w'\lambda} \cdot 2^{a+b}w'\lambda\left(\frac{\lambda}{n}\right)^{a+b-1} \leq 6\mu^4 2^\mu \frac{1}{6\mu^5 60^\mu} \leq 1$. So in total, the difference between the maximum and minimum values of $\sum_i \bar{\mathsf{p}}_1^a(i)\bar{\mathsf{p}}_2^b(i)\bar{k}^{a+b}$ is within $1 + |L^\mu_{a,b}|_1 \leq 1 + \mu^2\mu^a(2\mu)^b \leq (3\mu)^{a+b+2}$. We are interested in $\bar{k}$-based moments; since

[6] Note that here again we use vector multiplication where the vectors are indexed by pairs of numbers.

[7] $\lfloor v \rfloor$, where v is a vector, denotes the vector whose components are the floors of the components of v. Also, 1 as a vector denotes the vector with every entry equal 1.

[8] We use here $(\frac{\alpha}{\alpha+\beta})^t = (1 - \frac{\beta}{\alpha+\beta})^t \geq (1 - \frac{\beta}{\alpha})^t \geq 1 - t\frac{\beta}{\alpha}$.

$\frac{\bar{k}}{\lambda} = \frac{6kw\mu^5 60^\mu}{kw' \cdot 100 \cdot 2^{21\sqrt{\log n}}} \leq \frac{\mu^5 60^\mu}{2 \cdot 2^{21\mu}} \leq \frac{1}{20000\mu^4}$, the bound on the variation of $\sum_i \bar{\mathsf{p}}_1^a(i)\bar{\mathsf{p}}_2^b(i)\bar{k}^{a+b}$ is at most $(3\mu)^{a+b+2}\left(\frac{1}{20000\mu^4}\right)^{a+b} \leq \frac{1}{6000\mu^2}$ for $a+b \geq 2$.

Finally, for the last claim, note that the largest weight in p is bounded by $\frac{2\mu}{\lambda} < \frac{12\mu^6 2^{6\mu}}{kw'} < \frac{.84\mu^6}{2^{15\mu}} \cdot \frac{100 \cdot 2^{21\mu}}{kw} < \frac{1}{2^{14\mu}} 1/\bar{k}$. ■

Theorem 8 and Lemma 9 together implies the following.

Lemma 11 *For constant $0 < \alpha < 1$ and constant $c > 1$, any tester that distinguishes distribution pairs on $[n]$ that are in $P_{1/2}$ from pairs that are $\frac{3}{8}$-far from $P_{1/4}$ requires making $n^{1-o(1)}$ samples.*

Proof Consider the following two pairs of distributions $\mathsf{p} = (\mathsf{p}_1, \mathsf{p}_1)$ and $\mathsf{q} = (\mathsf{p}_2, \mathsf{p}_3)$ where p_1 is the distribution that is uniform on $[n]$, p_2 is the distribution that is uniform on $\{1, \ldots, n/2\}$ and zero elsewhere, and p_3 is the distribution that is uniform on $\{n/2+1, \ldots, n\}$ and zero elsewhere. Clearly, $\mathsf{p} \in P_{1/2}$. But q is $\frac{1}{2}$-far from $P_{1/4}$ because the closest distribution to q in $P_{1/4}$ is the pair $(\mathsf{p}_4, \mathsf{p}_3)$ where p_4 is the distribution that has weight $\frac{3}{4}$ uniformly on $\{1, \ldots, n/2\}$ and weight $\frac{1}{4}$ uniformly on $\{n/2+1, \ldots, n\}$ and this pair of distributions is $\frac{1}{2}$ far from q.

Distributions p_1, p_2, and p_3 all assign each element of $[n]$ weight at most $\frac{2}{n}$. Therefore, we can apply Lemma 9 with $k = n/2$ and $w = 1/8$, transforming the distribution pairs p and q to $\bar{\mathsf{p}}$ and $\bar{\mathsf{q}}$ respectively. From the lemma, $\bar{k} = \frac{kw}{100 \cdot 2^{21\sqrt{\log n}}}$. By (1) in Lemma 9, $\bar{\mathsf{p}} \in P_{1/2}$. By (2), $\bar{\mathsf{q}}$ is 3/8-far from $P_{1/4}$. We know from (3) of Lemma 9 that $|m_{\bar{k},\bar{\mathsf{p}}}(a,b) - m_{\bar{k},\bar{\mathsf{q}}}(a,b)| \leq \frac{1}{3000\log n}$. Then, setting $\rho = \frac{1}{2^{14\sqrt{\log n}}}$ and applying Theorem 8 implies that any tester that distinguishes between $P_{1/2}$ and 3/8-far from $P_{1/4}$ with probability at least 2/3 requires $\bar{k} = n^{1-o(1)}$ many samples. ■

Now, we are nearly done. For $\epsilon < 1/6$, Lemma 7 shows that an ϵ-tester for monotonicity with respect to M_n making $q(n, \epsilon)$ queries would result in an algorithm making $q(n, \epsilon/c)$ queries that would distinguish between instances in $P_{1/2}$ and instances ϵ-far from $P_{1/4}$ where $c > 1$ is an absolute constant. But since Lemma 11 rules out testers with $n^{1-o(1)}$ queries for distinguishing between being in $P_{1/2}$ and being 3/8-far from $P_{1/4}$, there must be no ϵ-tester for monotonicity with respect to M_n using less than $n^{1-o(1)}$ samples when $\epsilon < 1/6$. ■

2.2 Applications of the lower bound for the matching

The lower bound for the matching can be used as a building block for showing lower bounds for several other natural posets. To do so, the following straightforward lemma will be useful. An *up-set* in a poset $P = (V, \preceq)$ is a subset $U \subseteq V$ which is monotone nondecreasing. That is, if $u \in U, v \in V$ and $u \preceq v$, then $v \in U$. An up-set itself is a poset with the ordering given by $\preceq$. The next lemma shows that if Q is an up-set of P, then testing monotonicity with respect to Q is not harder than testing it with respect to P.

Lemma 12 *Suppose that a poset P with n elements contains an up-set Q. Then, testing monotonicity with respect to Q reduces to testing monotonicity with respect to P.*

Proof Any distribution p on Q can be viewed as a distribution on P, by setting $\mathsf{q}(x) = 0$ for all $x \in P \setminus Q$. We now show that the distance to monotonicity is unchanged.

Suppose that the distance of p from being monotone as a distribution on Q is ϵ, and that p' is a monotone distribution that is ϵ-close to p. Then p' can be extended to P just as p was extended, and it will still be monotone because Q is an up-set, showing that the distance to monotonicity over P is no more than ϵ.

Now suppose that the distance of p from being monotone as a distribution on P is $\delta \leq \epsilon$, and let $\bar{\mathsf{p}}$ be the distribution witnessing this. We now construct a distribution p' over Q. Let $\alpha = \sum_{x \in P \setminus Q} \bar{\mathsf{p}}$, and first construct $\tilde{\mathsf{p}}$ as the vector that is the truncation of p' to Q. If $\alpha > 0$ then this is not a probability vector, but clearly it is monotone and its L_1 distance from p is $\delta - \alpha$.

To finish the construction, we take a top-most element z of Q, that is an element of Q for which $z \preceq y$ implies $z = y$. Clearly $\tilde{\mathsf{p}}(z) \leq 1 - \alpha$, so to construct p' from $\tilde{\mathsf{p}}$ we just increase the value on z by α. This is now a monotone distribution, and by the triangle inequality its L_1 distance from p is not more than δ, showing that $\epsilon = \delta$. ■

Lemma 12 already shows that testing monotonicity with respect to a poset consisting of a linear number of disjoint chains requires $n^{1-o(1)}$ samples. The next corollary substantially generalizes the class of posets to which the nearly linear lower bounds apply. An *outward-directed star of degree d* refers to a directed graph with vertex set $c, v_1, \ldots, v_d$ and edge set $\{(c, v_i) : i \in [d]\}$.

Corollary 13 *Suppose that a poset P on n elements contains an up-set Q, which consists of $n^{1-o(1)}$ disjoint outward-directed stars of constant maximum degree. Then, testing monotonicity with respect to P requires $n^{1-o(1)}$ samples.*

Proof By Lemma 12, we only need to show the nearly linear lower bound for testing monotonicity with respect to Q. We do so by providing a reduction from testing monotonicity with respect to the matching poset M_r, where $r = n^{1-o(1)}$ is the number of disjoint outward-directed stars of constant degree that Q is composed of, and then applying Theorem 5.

Suppose that we have a distribution p on M_r. We arbitrarily map each edge in M_r to a distinct star in Q. Now, for an edge (x, y) in M_r mapped to the star with edges $\{(c, v_1), \ldots, (c, v_d)\}$ with $d \geq 1$, let $\mathsf{q}(c) = \mathsf{p}(x)/d, \mathsf{q}(v_1) = \mathsf{p}(y)/d$, and for all $i \in [2, d]$, let $\mathsf{q}(v_i) = (\mathsf{p}(x) + \mathsf{p}(y))/d$. q as defined is clearly a probability distribution on Q. If p is monotone, then q is also monotone. On the other hand, if p is ϵ-far from monotone with respect to M_r, then q is ϵ/d^*-far from monotone with respect to Q, where d^* is the maximum degree of a star in Q. This is because in any star with edges $\{(c, v_1), \ldots, (c, v_d)\}$, the only edge that could have monotonicity violated by q is (c, v_1) and the closest monotone distribution to q will not change the values of $\mathsf{q}(v_2), \ldots, \mathsf{q}(v_d)$. Furthermore, observe that given the ability to sample from p, we can generate a sample from q as follows. For an edge (x, y) in M_r mapped to the star $\{(c, v_i) : i \in [d]\}$ in Q, if p generates x, choose uniformly at random among the vertices $\{c, v_2, v_3, \ldots, v_d\}$ and if p generates y, choose uniformly at random among the vertices $\{v_1, v_2, \ldots, v_d\}$. ■

Corollary 13 implies, for example, that testing monotonicity with respect to the outward directed binary tree, or the *fence poset* (given by $x_1 \prec x_2 \succ x_3 \prec x_4 \succ x_5 \prec \cdots \succ x_n$), or, in fact, any poset described by a connected bipartite graph with bounded degree and all edges directed left to right, requires $n^{1-o(1)}$ samples. Regarding the last class of posets described, note that it is perhaps surprising that other structural properties of the bipartite graph (such as expansion) do not play any role at all in determining the sample complexity.

2.3 Testing Monotonicity with Respect to a Poset Containing a Large Matching

The lower bounds from the previous section do not apply when the poset contains long chains. Our next result shows that for such posets, $\Omega(\sqrt{n})$ samples are necessary.

Theorem 14 *If a poset P contains a matching of size $\Omega(n)$ in its transitive closure, then any monotonicity tester with respect to P requires $\Omega(\sqrt{n})$ samples.*

Proof We show two distributions, D_P and D_N, on positive (distributions monotone with respect to P) and negative (distributions ϵ-far from being monotone with respect to P for a constant ϵ) inputs respectively, such that any tester making $o(\sqrt{n})$ samples cannot distinguish, with high probability, between the case where the input is drawn from D_P and the case where it is drawn from D_N. We let D_P be always the input which is the uniform distribution over P (this is clearly a monotone distribution over P). To define D_N, we use the following lemma:

Lemma 15 *A uniformly chosen random boolean function $g : P \to \{0, 1\}$ is, with high probability, $\Omega(1)$-far (in L_1) from being monotone with respect to P, and furthermore has $\Omega(n)$ violated edges within a fixed in advance matching M of size $\Omega(n)$ in P.*

Proof Let M be the matching of size cn contained in the transitive closure of P. A random function g violates each edge of the matching with probability $1/4$, and so a uniformly chosen random function g has at least $cn/10$ edges violated with high probability, by Chernoff bounds. Since these edges are disjoint, g is also $c/10$-far from monotone with high probability. ■

D_N will be the distribution space chosen at random as follows. First, choose a uniformly random boolean function $g : P \to \{0, 1\}$. Now define D_N to be the distribution obtained, when with probability $1/3$, one of the zeros of g is uniformly chosen, and with probability $2/3$, one of the ones of g is uniformly chosen.

Note that if g has $\Omega(n)$ violated edges in M and has between $5n/12$ and $7n/12$ zeros (both of which happen with high probability), then the resulting D_N is indeed $\Omega(1)$-far (in L_1 distance) from being a monotone distribution. This is because such a number of zeros implies that for every violated edge of M we must change the distribution by at least $\frac{2}{3} \cdot \frac{12}{7n} - \frac{1}{3} \cdot \frac{12}{5n} = \frac{12}{35n}$.

To finish the proof, note that as long as the samples provided to the algorithm contain no collision (duplicate element), there is no way to distinguish D_P from a randomly chosen D_N. This is since, over both D_P and D_N, the distribution of the sequence of samples $x_1, \ldots, x_q$ conditioned on the event of having no collision is identical to a uniformly random choice of a non-repetitive sequence of q elements from P.

Now let α be the probability that the algorithm accepts a sample sequence chosen by a uniformly random choice of a non-repetitive sequence of q elements from P. If we are only allowed $q = o(\sqrt{n})$ samples, then the probability for a collision in our sample sequence is $o(1)$, and therefor over both D_P and D_N the algorithm will accept with probability $\alpha \pm o(1)$. Hence no algorithm can distinguish D_P from D_N using this many queries. ■

3 Monotonicity testers via path decomposition

In [BKR04], Batu, Kumar and Rubinfeld give a testing algorithm that shows that in the case of the total order on a domain of size n, the lower bound of $\Omega(\sqrt{n})$ samples from Theorem 14 is indeed tight to within polylogarithmic factors. They also consider distributions over the d-*dimensional grid poset*, which is the set $[m]^d$ ordered according to the dominance order, i.e., the relation $(x_1, \ldots, x_d) \preceq (y_1, \ldots, y_d)$ if and only if $x_i \leq y_i$ for all $i \in [d]$. Our next result gives a general bound that applies to any poset with a known chain decomposition. We then use this to settle the conjecture in [BKR04] regarding the sample complexity for monotonicity testing with respect to grid posets $[m]^d$, as well as to give sublinear sample complexity testers for distributions over other posets such as the Boolean hypercube.

Theorem 16 *Given a poset P that can be decomposed into a union of w disjoint chains of length at most c, there exists an ϵ-tester for monotonicity with respect to P with sample complexity $\tilde{O}(w\sqrt{c}\,\mathrm{poly}(1/\epsilon))$.*

Proof We use as a blackbox the following result from [BKR04].

Theorem 17 (Theorem 10 in [BKR04]) *There exists a randomized algorithm $\mathsf{ChainPartition}$ that, given a totally ordered set $\mathcal{L}$ of size n, parameters $\epsilon, \delta \in (0,1)$ and a random sample S of size $\Omega(\epsilon^{-4}\sqrt{n}\log n \log 1/\delta)$ from a probability distribution p on $\mathcal{L}$, acts as follows:*

- *With probability at least $1-\delta$, $\mathsf{ChainPartition}$ outputs either **FAIL** or a distribution $\mathsf{q} = \mathsf{ChainPartition}(\mathcal{L}, \epsilon, \delta, S)$ on $\mathcal{L}$ such that $\|\mathsf{q}-\mathsf{p}\|_1 < \epsilon$ (probabilities are taken over the internal coin tosses of $\mathsf{ChainPartition}$ and the guaranteed randomness of S as a sample taken from p).*
- *In particular, if p is ϵ-far from being monotone with respect to $\mathcal{L}$, then $\mathsf{ChainPartition}$ outputs **FAIL** with probability at least $1-\delta$.*
- *If p is monotone with respect to $\mathcal{L}$, then $\mathsf{ChainPartition}$ does not output **FAIL** with probability at least $1-\delta$.*

Our tester works as specified below. For a sample set S from a domain D and a subset $R \subseteq D$, we denote by $S|_R$ the set of samples that lie in R.

1. Set $m = \Theta(w\sqrt{c}\ \log(w)\ \log(c)\ \mathrm{poly}(1/\epsilon))$, and $\mu = \frac{\epsilon m}{20w}$ to be the sample size required by $\mathsf{ChainPartition}$ for $|\mathcal{L}| = w$, $\epsilon/4$ and $\delta = \frac{w}{200}$
2. Draw m samples from p. Call the sample sequence S.
3. For each chain $\mathcal{C}$ of the chain decomposition do:
   (a) Let $S_{\mathcal{C}}$ denote the subsequence of S consisting only of members of $\mathcal{C}$. If $|S_{\mathcal{C}}| < \mu$, let $\mathsf{q}_{\mathcal{C}}$ be the uniform distribution on the vertices of $\mathcal{C}$.
   (b) Otherwise, run the algorithm $\mathsf{ChainPartition}(\mathcal{C}, \epsilon/4, \delta, S_{\mathcal{C}})$, and output **FAIL** and terminate if it fails. Otherwise, let $\mathsf{q}_{\mathcal{C}}$ be the conditional distribution on $\mathcal{C}$ output by the algorithm.
4. Define a distribution $\tilde{\mathsf{q}}$ on P by setting the weight of a vertex v on a chain $\mathcal{C}$ to $\mathsf{q}_{\mathcal{C}}(v)\frac{S_{\mathcal{C}}}{m}$.
5. Output **PASS** if $\tilde{\mathsf{q}}$ is $\epsilon/2$-close to monotone with respect to P, and output **FAIL** otherwise.

The sample complexity claim is immediate. It remains to show completeness and soundness.

Lemma 18 *If p is a monotone distribution on P, then the above algorithm outputs **PASS** with probability at least $2/3$.*

Proof Call a chain $\mathcal{C}$ *light* if $\sum_{v\in\mathcal{C}} \mathsf{p}(v) < \frac{\epsilon}{10w}$, and *heavy* otherwise. We claim first that there are at least μ samples in S from every heavy chain, with constant probability. To see this, note that the expected number of samples in S from some given heavy chain $\mathcal{C}$ is at least $\frac{\epsilon m}{10w}$. Using the Chernoff bound, the probability that $|S_{\mathcal{C}}| < \mu = \frac{\epsilon m}{20w}$ is at most $\exp(-\Omega(\log(w)\ \sqrt{c}\ \log(c)\ \mathrm{poly}(1/\epsilon))) < \frac{1}{100w}$. By the union bound, then, with probability at least 0.99, each heavy chain is hit at least $\mu = \Omega(\sqrt{c}\ \log(c)\ \log(w)\ \mathrm{poly}(1/\epsilon))$ many times by the samples in S.

If p is monotone, then it is monotone on each chain. Now, if each heavy chain $\mathcal{C}$ is sampled $\Omega(\sqrt{c}\ \log(c)\ \log(w)\ \mathrm{poly}(1/\epsilon))$ many times, one can apply Theorem 17 to say that, with probability at least $1-\frac{1}{100w}$, the algorithm $\mathsf{ChainPartition}$ finds a distribution $\mathsf{q}_{\mathcal{C}}$ that is within $\epsilon/4$ in L_1-distance of $\mathsf{p}|_{\mathcal{C}}$ (the conditional distribution of p on the vertices of $\mathcal{C}$). Also, for each light chain $\mathcal{C}$, with probability at least $1 - \frac{1}{100w}$, $\frac{|S_{\mathcal{C}}|}{m} < \frac{\epsilon}{5w}$ by another application of the Chernoff bound. So, by the union bound and the triangle inequality, with probability at least 0.98, the distribution $\tilde{\mathsf{q}}$ is within $\epsilon/2$ in L_1-distance of p.

Taking the union bound with the event that the heavy chains are sampled sufficiently many times, we see that the algorithm outputs **PASS** in the last step, with probability at least $2/3$. ■

Lemma 19 *If p is a distribution on P such that the algorithm outputs **PASS** with probability at least $\frac{1}{3}$, then p is ϵ-close to a monotone distribution on P.*

Proof Let X be the event that there exists a chain $\mathcal{C}$ sampled at least μ times by the algorithm such that the distribution $\mathsf{q}_{\mathcal{C}}$ output by $\mathsf{ChainPartition}$ is $\epsilon/4$-far from $\mathsf{p}|_{\mathcal{C}}$. Let Y be the event that there exists a chain $\mathcal{D}$ sampled less than μ times by the algorithm such that $\sum_{v\in\mathcal{D}} \mathsf{p}(v) > \frac{3\epsilon}{20w}$. Let us upper-bound $\Pr[X|\text{algorithm outputs } \textbf{PASS}]$ and $\Pr[Y]$, where the probabilities are over the randomness of the algorithm and the sample. For the first, notice that since the event is conditioned on the algorithm passing, step (3b) never fails; hence, using Theorem 17 and the union bound, $\Pr[X|\text{algorithm outputs } \mathbf{PASS}] \le w \cdot \frac{1}{100w} = 0.01$. For the second, observe that the probability that a chain $\mathcal{D}$ is hit less than μ times by S while having over $\frac{3\epsilon}{20w}$ weight under p is at most $\frac{1}{100w}$ by Chernoff bounds; so, by the union bound, $\Pr[Y] < 0.01$.

Now, using the union bound again, we have:

$$\begin{aligned}
&\Pr[\text{Algorithm outputs } \mathbf{PASS} \wedge \neg X \wedge \neg Y] \\
&\quad \ge \Pr[\text{Algorithm outputs } \mathbf{PASS}] - \Pr[X \wedge \text{Algorithm outputs } \mathbf{PASS}] - \Pr[Y] \\
&\quad \ge \Pr[\text{Algorithm outputs } \mathbf{PASS}] - \Pr[X|\text{Algorithm outputs } \mathbf{PASS}] - \Pr[Y] \\
&\quad \ge \frac{1}{3} - 0.01 - 0.01 > 0 \qquad (1)
\end{aligned}$$

So, there exists a distribution $\tilde{\mathsf{q}}$ on P such that $\tilde{\mathsf{q}}$ is $\epsilon/2$-close to a monotone distribution (because the algorithm accepts). Furthermore, $\|\tilde{\mathsf{q}} - \mathsf{p}\|_1 \le \epsilon/2$, because $\|\tilde{\mathsf{q}} - \mathsf{p}\|_1 \le \frac{\epsilon}{4} + w(\frac{\epsilon}{10w} + \frac{3\epsilon}{20w})$, where the first term is the contribution of chains with at least μ samples and the second term is from the rest of the chains. By the triangle inequality, p is ϵ-close to a monotone distribution on P. ■

The above concludes the proof of Theorem 16. ■

The following corollary is immediate:

Corollary 20 *For any finite poset P on n elements with width w, there exists a monotonicity tester with respect to P with sample complexity $\tilde{O}(w\sqrt{n}\ \mathrm{poly}(1/\epsilon)$.*

Proof Dilworth's theorem states that if w is the width of P (i.e., size of the longest antichain), then P can be decomposed into a union of w disjoint chains. Each chain is clearly of length at most n. ■

The next corollary gives a sublinear sample tester for monotonicity with respect to the d-dimensional cube, resolving the conjecture in [BKR04]. In the following, note that the size of the domain is $n = m^d$.

Corollary 21 *Let $C_{m,d}$ be the d-dimensional grid poset with elements $[m]^d$. Then, there exists an ϵ-monotonicity tester with respect to C_d with sample complexity $\tilde{O}(m^{d-1/2}\mathrm{poly}(1/\epsilon))$, where the asymptotic notation refers to a fixed d where m is growing.*

Proof Consider the following chain decomposition of $C_{m,d}$. For $\sigma \in [m]^{d-1}$, let $C_\sigma = \{(\sigma_1, \dots, \sigma_{d-1}, i) : i \in [m]\}$. It is clear that each C_σ is a chain and that they partition $C_{m,d}$. Moreover, each chain is of length m. So, applying Theorem 16 yields the sample complexity bound. ■

The next result achieves the first sublinear time monotonicity tester for the Boolean hypercube. In the following note that the size of the domain is $n = 2^d$.

Corollary 22 *Let H_d denote the poset on $\{0,1\}^d$ induced by the usual subset order. Then, there exists an ϵ-monotonicity tester with respect to H_d with sample complexity $\tilde{O}\left(\frac{2^d}{(d/\log d)^{1/4}}\mathrm{poly}(1/\epsilon)\right)$.*

Proof We use the main result of [HLST03] that states that H_d can be decomposed into $\binom{d}{\lfloor d/2 \rfloor}$ chains, each of size $O(\sqrt{d \log d})$. Using the Stirling approximation, and applying Theorem 16, immediately gives the desired bound. ■

3.1 About the optimality of the tester.

Theorem 16 is not tight. For example, consider the outward-directed star graph on n vertices. There is a simple $O(1/\epsilon)$ sample monotonicity tester with respect to this poset. To see this, observe that any monotone distribution places weight at most $1/n$ at the center vertex, and any distribution ϵ-far from monotone places weight at least ϵ on the center vertex. Hence, checking whether the center vertex appears in a random sample of $O(1/\epsilon)$ samples suffices to distinguish between the two cases, with high probability. However, for the chain decomposition of the outward-directed star, the antichain size is linear in n, and hence, the resulting tester from Theorem 16 is far from optimal. Another example of a poset with a large chain-antichain decomposition but requiring only a small number of samples for monotonicity testing is given by the inward-directed star.

Theorem 23 *There exists a monotonicity tester with respect to the inward-directed star with sample complexity $\tilde{O}(1/\epsilon^2)$.*

Proof Here is our tester. Given an input distribution p on the inward-directed star:

1. Assume that ϵ is smaller than some global constant $\hat{\epsilon}$. Otherwise, perform the rest of the algorithm for $\hat{\epsilon}$ instead of ϵ.
2. Sample $m = O(1/\epsilon^2 \log 1/\epsilon)$ samples from p.
3. Let u denote the center vertex of the poset, and for every v let $\mathit{count}(v)$ denote the number of times that v appeared in the sample sequence.
4. Accept if and only if $\mathit{count}(u) \geq \max_{v \neq u} \mathit{count}(v) - \epsilon m/4$.

For the analysis, let u denote the center vertex of the star, while $v_1, \ldots, v_n$ denote the outside vertices. For a vertex x, we will call the fraction of times that x is sampled among the m samples as the algorithm's estimate of the probability weight of x.

We first observe that with probability at least $\frac{5}{6}$, for each vertex x whose actual weight is at least ϵ^5, the algorithm estimates the probability weight of x to within an additive error of $\epsilon/8$. This is because for any given vertex, the probability that the algorithm's estimate of its weight differs from its actual weight by more than $\epsilon/8$ is at most $O(\epsilon^5)$ by Chernoff bounds. Since there are at most $O(1/\epsilon^5)$ vertices with such weights, we can ensure that the probability of being off by more than $\epsilon/8$ in the estimate of any such vertex is at most $\frac{5}{6}$.

After we set the coefficient in the number of samples to be used as per the above paragraph, we take care of setting $\hat{\epsilon}$. Let U be the set of vertices whose weight is at most ϵ^5, and $m = C \cdot (1/\epsilon^2 \log 1/\epsilon)$ be the number of samples (with C the constant fixed by the above). The probability of any of the vertices from U to appear more than once in the sample is bounded by $\sum_{x \in U} \binom{m}{2} (\mathsf{p}(x))^2 \leq \binom{m}{2} \max_{x \in U} \mathsf{p}(x) \leq \frac{1}{2} C^2 (1/\epsilon^2 \log 1/\epsilon)^2 \epsilon^5$. A proper choice of $\hat{\epsilon}$ makes this smaller than $\frac{1}{6}$, and so with proability at least $\frac{5}{6}$ none of the elements with weight less than ϵ^5 will appear in more than one sample. We also make sure that $\epsilon^5 \leq \epsilon/8$.

From now on we suppose that both of the above events occur (which they do with probability at least $\frac{2}{3}$). In particular they mean that no vertex received an estimation that is more than $\epsilon/8$ away from its actual weight in any direction.

Now, suppose that p is a monotone distribution. This means that the weight of u is at least the maximum weight of any other vertex in the poset. At the most, the algorithm has mis-estimated both the weight of u and the maximum weight of the other vertices by at most $\epsilon/8$. This means that the algorithm accepts, since the estimated weight of u will be not less than the maximum estimated weight of the other vertices minus $\epsilon/4$.

On the other hand, suppose now that p is ϵ-far from monotone. In particular this means that the weight of u is less than the maximum weight of the other vertices minus $\epsilon/2$, as otherwise we just decrease the weight of u by $\epsilon/2$ at the expense of the other vertices and obtain a monotone ϵ-close distribution. As both $count(u)$ and $\max_{v \neq u} count(v)$ reflect the actual weights with up on an $\epsilon/8$ error, the algorithm rejects. ■

4 Acknowledgements

We would like to thank Adi Akavia, Punyashloka Biswal, and Rocco Servedio for useful and encouraging discussions.

References

[AAK+07] Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing k-wise and almost k-wise independence. In *Proc. 39th Annual ACM Symposium on the Theory of Computing*, pages 496–505, 2007.

[ACS10] Michał Adamaszek, Artur Czumaj, and Christian Sohler. Testing monotone continuous distributions on high-dimensional real cubes. In *Proc. 23rd ACM-SIAM Symposium on Discrete Algorithms*, pages 56–65, 2010.

[AKNS99] Noga Alon, Michael Krivelevich, Ilan Newman, and Mario Szegedy. Regular languages are testable with a constant number of queries. In *Proc. 40th Annual IEEE Symposium on Foundations of Computer Science*, page 645, 1999.

[BDKR02] Tugkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. In *Proc. 34th Annual ACM Symposium on the Theory of Computing*, pages 678–687, 2002.

[BFF+01] Tugkan Batu, Lance Fortnow, Eldar Fischer, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 442–451, 2001.

[BFR+00] Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *Proc. 41st Annual IEEE Symposium on Foundations of Computer Science*, pages 259–269, 2000.

[BKR04] Tugkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proc. 36th Annual ACM Symposium on the Theory of Computing*, pages 381–390, 2004.

[BRW99] Tugkan Batu, Ronitt Rubinfeld, and Patrick White. Fast approximate pcps for multidimensional bin-packing problems. In *RANDOM-APPROX*, pages 245–256, 1999.

[BS07] Mickey Brautbar and Alex Samorodnitsky. Approximating entropy from sublinear samples. In *Proc. 18th ACM-SIAM Symposium on Discrete Algorithms*, pages 366–375, 2007.

[Dev91] Luc Devroye. Algorithms for generating discrete random variables with a given generating function or a given moment sequence. *SIAM Journal on Scientific and Statistical Computing*, 12(1):107–126, 1991.

[DGL+99] Yevgeniy Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. In *RANDOM-APPROX*, pages 97–108, 1999.

[EKK+00] Funda Ergun, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Mahesh Vishwanthan. Spot-checkers. *J. Comp. Sys. Sci.*, 60:717–751, 2000. Earlier version in STOC'96.

[Fis01] Eldar Fischer. The art of uninformed decisions: A primer to property testing. *The Computational Complexity Column of the Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001.

[FLN+02] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proc. 34th Annual ACM Symposium on the Theory of Computing*, pages 474–483, 2002.

[FN01] Eldar Fischer and Ilan Newman. Testing of matrix properties. In *Proc. 33rd Annual ACM Symposium on the Theory of Computing*, pages 286–295, 2001.

[GGL+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

[GMV06] Sudipto Guha, Andrew McGregor, and Suresh Venkatasubramanian. Streaming and sublinear approximation of entropy and information distances. In *Proc. 17th ACM-SIAM Symposium on Discrete Algorithms*, pages 733–742, 2006.

[GR00] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity (ECCC)*, TR00-020, 2000.

[HLST03] Tim Hsu, Mark J. Logan, Shahriar Shahriari, and Christopher Towse. Partitioning the boolean lattice into a minimal number of chains of relatively uniform size. *European Journal of Combinatorics*, 24(2):219 – 228, 2003.

[Kli67] Allen Klinger. The Vandermonde Matrix. *The American Mathematical Monthly*, 74(5):571–574, 1967.

[Pan08] Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

[RRSS07] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 559–569, 2007.

[RS09] Ronitt Rubinfeld and Rocco A. Servedio. Testing monotone high-dimensional distributions. *Random Struct. Algorithms*, 34(1):24–44, 2009.

[Val07] Paul Valiant. Testing symmetric properties of distributions. *Electronic Colloquium on Computational Complexity (ECCC)*, TR07-135, 2007.

[Val08] Paul Valiant. Testing symmetric properties of distributions. In *Proc. 40th Annual ACM Symposium on the Theory of Computing*, pages 383–392, 2008.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de