New Lower Bounds for Testing Monotonicity and Log Concavity of Distributions

Yuqian Cheng
chengyuqian6@gmail.com

Daniel M. Kane[†]
UC San-Diego
dakane@ucsd.edu

Zhicheng Zheng
New York University
zz4230@nyu.edu

August 2, 2023

1 Introduction

1.1 Background

Given data from an unknown distribution, can one determine whether the underlying distribution has certain properties or not? In order to make this determination, how much data would be needed? These are the critical questions in the field of statistical hypothesis testing (see [7]). Some of these questions, particularly when dealing with discrete distributions, have over the past few decades drawn the interest of computer scientists under the heading of distribution testing (see [8]).

In order to make such a determination possible, distribution testing algorithms are usually asked to distinguish between the cases where the unknown distribution p either has the property $\mathcal{P}$ or is far from (usually in L^1 distance) any distribution with this property. The goal is usually to do this with as few samples as possible (ideally to within a constant factor of the information-theoretic limits) and in a computationally efficient manner. While many of the most basic questions such as testing for uniformity, identity, closeness and independence have all been resolved, many more complicated properties still have wide gaps between the best known algorithms and the best lower bounds. For a survey of progress in this field see [4] for a recent survey of the area.

In this paper, we develop a new lower bound technique for testing properties that are defined by inequalities between the probability masses of individual bins. In particular, this technique is used to produce new lower bounds for testing monotonicity and log-concavity, the latter of which matches known upper bounds to within polylog factors.

1.2 Notation

We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. As we will usually be dealing in this paper with discrete distributions, for a distribution p on a discrete set S and an element $i \in S$, we let p_i denote the probability that p assigns to the element i. The distance considered in this paper will usually be the total variational

[*] Authors are listed in randomized order.

[†] Supported by NSF Medium Award CCF-2107547, NSF Award CCF-1553288 (CAREER), a Sloan Research Fellowship, and a grant from CasperLabs.

distance denoted by d_{TV}, defined as $d_{TV}(p,q) := \frac{1}{2}||p-q||_1$. In particular, if p and q are distributions defined over the same discrete set S, $d_{TV}(p,q) = \frac{1}{2}\sum_{i\in S}|p_i - q_i|$.

By an *ensemble* we will typically mean a probability distribution over probability distributions on some fixed set S.

1.3 Our Results

Our main applications have to do with the monotonicity test and log-concavity testing problems, so we begin by defining our terms.

Definition 1. *We say that a distribution p on $[n]$ is* monotone *(decreasing) if $p_i \geq p_j$ for all $i < j$.*

In fact, we will also deal with monotone distributions defined over higher dimensional cubes. To make sense of this we first need to define a partial order on $[n]^d$:

Definition 2. *For $\boldsymbol{i}, \boldsymbol{j} \in [n]^d$, we say that $\boldsymbol{i} < \boldsymbol{j}$ if $\boldsymbol{i}_a < \boldsymbol{j}_a$ for all $1 \leq a \leq d$.*

We say that a distribution p on $[n]^d$ is monotone *if $p_{\boldsymbol{i}} \geq p_{\boldsymbol{j}}$ whenever $\boldsymbol{i} < \boldsymbol{j}$.*

We also define log-concavity as follows:

Definition 3. *A distribution p on $[n]$ is* log-concave *if its support is a contiguous interval and its density function, p_i, satisfies that $p_i^2 \geq p_{i+1}p_{i-1}$ for all $2 \leq i \leq n-1$.*

Our main results are to prove new lower bounds for multidimensional monotonicity testing and log concavity testing. In particular, we show that algorithms that can reliably distinguish between a distribution having the desired property (monotonicity or log-concavity) and being ϵ-far in total variation distance from any such distribution, must make use of a relatively large number of samples.

The result for multidimensional monotonicity testing is stated as below.

Theorem 4. *For $\epsilon > 0$ sufficiently small, any algorithm that can distinguish whether the distribution over $[n]^d$ is monotone (for some n at least a sufficiently large multiple of $d\log(1/\epsilon)$) or ϵ far from monotone with probability over $\frac{2}{3}$ requires at least $N = 2^{-O(d)}d^{-d}\epsilon^{-2}\log^{-7}(1/\epsilon)\min(n, d\epsilon^{-1}\log^{-3}(1/\epsilon))^d$ samples.*

We get the following lower bound of log concavity testing of distribution over $[n]$.

Theorem 5. *For $\epsilon > 0$ sufficiently small, let p be distribution over $[n]$ where n is at least a sufficiently large multiple of $\log(1/\epsilon)$. Any algorithm that can distinguish whether the distribution is log concave or ϵ far from log concave with probability over $\frac{2}{3}$ requires at least $N = \Omega(\log^{-7}(1/\epsilon)\epsilon^{-2}\min(n, \epsilon^{-1/2}\log^{-3/2}(1/\epsilon)))$ samples.*

1.4 Prior and Related Works

The problem of monotonicity distribution testing was first considered by [2] who developed a tester with sample complexity $O(\frac{\sqrt{n}\log n}{\epsilon^4})$ for distributions over $[n]$. Then the result was generalized to give a tester with sample complexity $\widetilde{O}(n^{d-\frac{1}{2}}\text{poly}(\frac{1}{\epsilon}))$ for distributions over $[n]^d$ by [3]. The best currently known result is by [1], $O(\frac{n^{\frac{d}{2}}}{\epsilon^2} + (\frac{d\log n}{\epsilon^2})^d\frac{1}{\epsilon^2})$ for testing monotonicity of a distribution over $[n]^d$. On the other hand, the best lower bounds to date for this problem come from the lower bounds for uniformity testing giving a sample complexity of $\Omega(\sqrt{n^d}/\epsilon^2)$. While this shows that the algorithm of [1] is asymptotically optimal for n much larger than $1/\epsilon$, it leaves a pretty substantial gap when ϵ is small.

For testing log concavity of a distribution over $[n]$, [1] provides the first known tester for the low sample regime of testing log concavity, which requires $O(\frac{\sqrt{n}}{\epsilon^2}+\frac{1}{\epsilon^5})$ samples. Then [5] gives an improved algorithm with sample complexity $O(\frac{\sqrt{n}}{\epsilon^{\frac{7}{2}}})$. The latest result for sample complexity of log concavity testing lies in [6], $O(\frac{\sqrt{n}}{\epsilon^2})+\widetilde{O}(\frac{1}{\epsilon^{\frac{5}{2}}})$ for testing a distribution over $[n]$. Once again, previously known lower bounds for this problem were somewhat lacking, consisting only of the $\Omega(\sqrt{n}/\epsilon^2)$ lower bound for uniformity testing. However, importantly, if one combines this with our lower bound, it nearly matches the best of the upper bound of [6] and the trivial $O(n/\epsilon^2)$ bound from learning p to error ϵ. Together these show that the optimal sample complexity for testing log-concavity is $\left(\frac{\sqrt{n}}{\epsilon^2}+\frac{\min(n,\epsilon^{-1/2})}{\epsilon^2}\right)$ up to polylogarithmic factors.

1.5 Our Techniques

Our techniques come from a general framework for obtaining lower bounds for testing properties defined by imposing inequalities on the individual bin probabilities. Our starting point is fairly standard for such lower bounds: we want to construct two ensembles of distributions over some support set S, D_{yes} and D_{no} where the distributions in D_{yes} have the property with high probability and distributions in D_{no} far from having the property with high probability. Therefore, if an algorithm can distinguish a distribution having the property or ϵ far from having the property through N samples with probability at least $\frac{2}{3}$, then it should be able to distinguish N samples from a distribution in D_{yes} and N samples from a distribution in D_{no} with probability at least $\frac{3}{5}$. However, we show that if a distribution p is randomly taken from either D_{yes} or D_{no}, then a small number N of samples from p will be insufficient to reliably determine which of the two ensembles it was sampled from. In particular, our goal will be to show that the two resulting distributions on S^N will be close in total variational distance, making such a determination information-theoretically impossible.

In order to construct these ensembles, a key insight is that we will need them to match moments. In particular, if n_i is the number of samples drawn from the i^{th} bin then the expectation of $\binom{n_i}{k}$ will be $\binom{N}{k}\mathbb{E}[p_i^k]$. If we want our ensembles to be nearly indistinguishable, it is thus a good idea to ensure that these moments- $\mathbb{E}_{p\sim D}[p_i^k]$- are the same for D_{yes} and D_{no} for all small (in our case at most logarithmic) values of k. This will typically ensure that D_{yes} and D_{no} are hard to distinguish.

On the other hand, if the property in question is defined by inequalities among the p_i (like $p_i \geq p_{i+1}$ for monotonicity or $p_i^2 \geq p_{i-1}p_{i+1}$ for log-concavity), then these kinds of inequalities are not defined by moments. In particular, our goal will be to find a moment matching pair of ensembles D_{yes} and D_{no} so that distributions from D_{yes} satisfy these inequalities, but distributions from D_{no} do not.

For the mechanics of this construction, we begin by constructing a pair of moment distributions over real numbers F_{yes} and F_{no} with $\mathbb{E}[F_{yes}^k]=\mathbb{E}[F_{no}^k]$ for all small natural numbers k. These will be used to modify the bin probabilities of a base distribution Q. In particular, in a sample from $D_{yes/no}$ the bin probability of a bin p_i will by $Q_i + A_iF_{yes/no}$ for some carefully chosen constants A_i and with the samples from $F_{yes/no}$ coupled in such a way to ensure that p remains properly normalized. On the other hand, we can construct our distributions $F_{yes/no}$ so that while F_{yes} is positive almost surely, F_{no} has a reasonable probability of being reasonably negative, which (assuming that Q was constructed carefully) will break the inequalities defining the property in question.

In Section 2, we will explain the generic version of the construction of these ensembles D_{yes} and D_{no} and in particular prove Proposition 9 to show that they are indistinguishable with a small number of samples. The next three sections will be applications. In particular, in Section 3, as a warmup we will prove a lower bound for one-dimensional monotonicity testing. In Section, 4 we will generalize this to a lower bound for multidimensional monotonicity testing. Finally, in Section 5 we will prove our lower bound for testing log-concavity.

2 Generic Lower Bound Construction

This section contains the key construction for our lower bound technique. In particular, we provide a general framework for producing two ensembles of distributions D_{yes} and D_{no} over some finite set S that are hard to distinguish using few samples. In particular, for a positive integer N, we consider the two distributions over S^N which we call D_{yes}^N and D_{no}^N given by taking a random distribution p from the ensemble $D_{yes/no}$ and then returning N i.i.d. samples from p. The key result here is Proposition 9, where we show that as long as N is not too large relative to the other parameters of the construction that $d_{TV}(D_{yes}^N, D_{no}^N)$ is small, thus implying that one cannot reliably determine whether p was taken from D_{yes} or D_{no} with only N samples.

2.1 Construction of D_{yes} and D_{no}

In this section, we describe the general procedure for construction ensembles D_{yes} and D_{no}. The basic idea is to start with a fixed distribution Q over S and to tweak it slightly. In order to ensure that these tweaks match moments and preserve indistinguishability, we will first need to find a pair of real-valued distributions F_{yes} and F_{no} that match their low order moments, and we will use the outputs of these distributions to tweak the bin probabilities of Q. In order to ensure that the resulting distribution remains properly normalized, we will pair up a number of the bins in S getting pairs $(j_1, k_1), (j_2, k_2), \ldots, (j_s, k_s)$ and ensure that any probability mass taken from j_i is added to k_i and visa versa. Finally, to decide the amount of mass to move we will sample δ_i proportional to $F_{yes/no}$ and our final distribution will have $p_{j_i} = Q_{j_i} + \delta_i$ and $p_{k_i} = Q_{k_i} - \delta_i$.

We begin by defining our distributions F_{yes} and F_{no}, for which we will need to have a few free parameters:

Definition 6. *Let m be an integer, A and g be real numbers and a be a uniformly random integer $\bmod m$, we define the probability distributions $F_{yes}^{A,g,m}$ and $F_{no}^{A,g,m}$ to be the distribution of $A(\cos(\frac{2\pi a}{m}) + g)$ and $A(\cos(\frac{2\pi(a+\frac{1}{2})}{m}) + g)$.*

The first critical property of these distributions is that they match their first $m-1$ moments, which we can prove by making use of the Chebyshev polynomials $T_m(\cos\theta) := \cos(m\theta)$ for $\theta \in [0, \pi]$.

Lemma 7. *For any positive integer k less than m,*

$$E_{\delta \in F_{yes}^{A,g,m}}[\delta^k] = E_{\delta \in F_{no}^{A,g,m}}[\delta^k].$$

Proof. Note that the roots of $T_m(x) + 1$ and $T_m(x) - 1$ are $\cos(\frac{2\pi(a+\frac{1}{2})}{m})_{0 \le a < m}$ and $\cos(\frac{2\pi a}{m})_{0 \le a < m}$ respectively. Since $T_m(x) + 1$ and $T_m(x) - 1$ only differ by a constant, all elementary symmetric polynomials of degree less than m of roots of one agree with the corresponding polynomials of roots of the other. By the fundamental theorem on symmetric polynomials, for roots of a polynomial r, $\sum_r r^k$ can be written in terms of elementary symmetric polynomials, where $\sum_r r^k$ is proportional to the kth moment of roots. Since the roots have $m-1$ identical elementary symmetric polynomials, we can conclude that they have $m-1$ matching moments. In particular, this means that

$$\mathbb{E}\left[\cos(\frac{2\pi(a+\frac{1}{2})}{m})\right] = \mathbb{E}\left[\cos(\frac{2\pi a}{m})\right].$$

Applying the linear transformation $x \to A(x+g)$, we note that the distributions $A(\cos(\frac{2\pi a}{m}) + g)$ and $A(\cos(\frac{2\pi(a+\frac{1}{2})}{m}) + g)$ must also have $m-1$ matching moments as

$$\mathbb{E}[(A(x+g))^k] = \sum_{k'=0}^{k} A^k \binom{k}{k'} g^{k-k'} \mathbb{E}[x^{k'}].$$

□

We are now ready to define D_{yes} and D_{no} below:

Definition 8. *Suppose that we have:*

- *A distribution Q over a finite set S.*
- *A set of disjoint pairs of elements of S: $(j_1, k_1), (j_2, k_2), \ldots, (j_s, k_s)$.*
- *A positive integer m.*
- *Two sequences of real numbers $(A_i)_{1\leq i\leq s}$ and $(g_i)_{1\leq i\leq s}$ so that $|A_i| \leq \min(\frac{Q_{j_i}}{1+|g_i|}, \frac{Q_{k_i}}{1+|g_i|})$ for all i.*

Given this, we define a pair of ensembles of distributions over S, D_{yes} and D_{no} as follows:

To select a distribution p from $D_{yes/no}$, we first select δ_i independently from $F_{yes/no}^{A_i,g_i,m}$ for each $1 \leq i \leq s$. For $a \in S$ with a not equal to any j_i or k_i, we let $p_a = Q_a$. Otherwise, we let $p_{j_i} = Q_{j_i} + \delta_i$ and $p_{k_i} = Q_{k_i} - \delta_i$.

Note that as $|A_i| \leq \min(\frac{Q_{j_i}}{1+|g_i|}, \frac{Q_{k_i}}{1+|g_i|})$, p_a in non-negative for all $a \in S$. In addition, for each i, $p_{j_i} + p_{k_i} = Q_{j_i} + Q_{k_i}$, from which it is not hard to see that $\sum_{a\in S} p_a$ is always 1. These observations confirm that p is in fact defines a probability distribution over S.

Another important remark is that conditioned on a sample from p landing in the ith pair of bins, (j_i, k_i), the probability of it landing in the j_ith bin depends only on the value of δ_i, and δ_is are independently sampled from $F_{yes/no}^{A_i,g_i,m}$. This is a crucial condition our key proposition needs.

For this construction, we are hoping to prove the following proposition,

Proposition 9. *Given $Q, (A_i)_{1\leq i\leq s}, (g_i)_{1\leq i\leq s}, m, (j_i, k_i)_{1\leq i\leq s}$ be as above and let D_{yes} and D_{no} be as in Definition 8. Assume furthermore, that m is at least a $C\log(s)$ for some sufficiently large constant C.*

For integers $N > \frac{6\log s}{\min_i(Q_{j_i}+Q_{k_i})}$, we define D_{yes}^N to be the distribution on S^N obtained by first taking a random distribution p from D_{yes} and then taking N independent samples from p, and define D_{no}^N similarly. Then letting $x_{max} = \max_{1\leq i\leq s}\left(\frac{|A_i|(1+|g_i|)}{\min(Q_{k_i},Q_{j_i})}\right)$ and $B = 2\max_{1\leq i\leq s}(Q_{j_i} + Q_{k_i})N$, then if $x_{max} < 1/10$, we have that $d_{TV}(D_{yes}^N, D_{no}^N)$ is at most

$$O(1/s) + m^4 s O(\sqrt{B\log(s)} + x_{max}B)(1 + x_{max})^{O(\sqrt{B\log(s)}+x_{max}B)} O(\sqrt{x_{max}^2 B\log(s)} + x_{max}^2 B)^m.$$

In our applications, we will take s on the order of $|S|$ and will use Q's which are not too far from uniform (and thus $\min_{a\in S} Q_a$ will be on the order of $1/s$). In order to ensure that D_{yes} and D_{no} perturb Q by at least ϵ in total variational distance, we will want x_{max} (which is essentially the largest relative perturbation of any bin probability) to be on the order of ϵ. Taking N on the order of s/ϵ^2 up to some polylog factors gives us B on the order of $1/\epsilon^2$.

From here we note that the

$$(1 + x_{max})^{O(x_{max}B+\sqrt{B\log(s)})}$$

term is

$$\exp(O(Bx_{max}^2 + \sqrt{Bx_{max}^2\log(s)})),$$

which is not too large. On the other hand, so long as we ensure that

$$(x_{max}\sqrt{B\log(s)} + Bx_{max}^2)$$

is less than a sufficiently small constant and keep m to be a large enough multiple of $\log(s)$, this term will dominate the things it is multiplied by, thus leaving us with a final bound that is quite small.

In particular, we have

Corollary 10. *In the notation of Proposition 9, if we have additionally that Bx_{max}^2 is at most a sufficiently small multiple of $1/\log(s)$, m is at least a sufficiently large multiple of $\log(s/(x_{max}\epsilon))$ and s is at least a sufficiently large constant, then*

$$d_{TV}(D_{yes}^N, D_{no}^N) < \frac{1}{100}.$$

Proof. Assume that for some A sufficiently large that $Bx_{max}^2 \log(s) < 1/A$ and that $m \geq A\log(s/(x_{max}\epsilon))$. Then the bound in Proposition 9 reduces to

$$O(1/s)+O(m^4 s/x_{max})(\sqrt{Bx_{max}^2\log(s)}+x_{max}^2 B)\exp(O(\sqrt{Bx_{max}^2\log(s)}+x_{max}^2 B))O(\sqrt{Bx_{max}^2\log(s)}+x_{max}^2 B)^m.$$

Noting that $(\sqrt{Bx_{max}^2\log(s)}+x_{max}^2 B) = O(1/A)$, this reduces to

$$O(1/s)+O(m^4 s/(Ax_{max}))\exp(O(1/A))O(1/A)^m.$$

In particular, if A is large enough the $O(1/A)^m$ term is at most $(1/2)^m$, which if m is a sufficiently large multiple of $\log(s/(x_{max}\epsilon))$ is at most $m^{-4}x_{max}/s^2$, which would make our final bound $O(1/s)$. If s is sufficiently large, this is less than 1/100. □

2.2 Comparison of Distributions of Number of Samples in Bins j_i and k_i

In our construction of D_{yes} and D_{no} we refer to the ith pair of bins as the pair $\{j_i, k_i\}$. As D_{yes} and D_{no} are essentially identical except in how they distribute probability mass between the ith pair of bins for various values of i, in order to show that they are hard to distinguish, it will be important for us to show that the distribution on the number of samples in these bins is close for D_{yes} and D_{no}. In particular, if we condition on the number of samples B_i that land in the ith pair of bins, and consider the probability that exactly ℓ_i samples lie in the first of this pair (i.e. j_i), we would like to show that this probability is similar for a random distribution from D_{yes} and a random distribution from D_{no}. In particular, we prove:

Lemma 11. *Let $A_i, g_i, m, Q, (j_i, k_i)$ be as in Proposition 9, and let $1 \leq i \leq s$, and let B_i and N be non-negative integers. Then we have that if N i.i.d. samples are taken from a probability distribution p taken from either D_{yes} or D_{no}, and consider this distribution conditioned on exactly B_i samples lying in the ith pair of bins. Let X_i^{yes} and X_i^{no} be the distributions on the number of samples drawn from the bin j_i in the case where p is taken from D_{yes} or D_{no} respectively. Then*

$$d_{TV}(X_i^{yes}, X_i^{no}) \leq O(1/s^2)+m^4 O(\sqrt{B_i\log(s)}+x_{max}B_i)(1+x_{max})^{O(\sqrt{B_i\log(s)}+x_{max}B_i)}O(\sqrt{x_{max}^2 B_i\log(s)}+x_{max}^2 B_i)^m.$$

Proof. We begin by proving this in the case where $Q_{j_i} = Q_{k_i}$ and will reduce to this case later.

The key observation here is that if we let $x = \frac{\delta_i}{Q_{j_i}}$ then a sample landing in the ith pair of bins will have a probability of

$$\frac{p_{j_i}}{p_{j_i}+p_{k_i}} = \frac{Q_{j_i}+\delta_i}{Q_{j_i}+Q_{k_i}} = \frac{1+x}{2}$$

of landing in bin j_i. Thus, conditioned on δ_i, $X_i^{yes/no}$ is distributed as the binomial distribution $\mathrm{Bin}(B_i, (1+x)/2)$. Therefore, the probabilities of X_i^{yes} and X_i^{no} being equal to ℓ are just

$$\mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[\Pr(\mathrm{Bin}(B_i,(1+x)/2)=\ell)] = 2^{-B_i}\binom{B_i}{\ell}\mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}\left[(1+x)^{\ell}(1-x)^{B_i-\ell}\right]$$

and

$$2^{-B_i}\binom{B_i}{\ell}\mathbb{E}_{x\sim F_{no}^{A_i,g_i,m}/Q_{j_i}}\left[(1+x)^{\ell}(1-x)^{B_i-\ell}\right],$$

respectively.

Our goal will be to show that (at least for ℓ close to $B_i/2$, which it will be with high probability) that these are close. The basic plan here is to approximate the term $(1+x)^\ell(1-x)^{B_i-\ell}$ by its Taylor series about $x=0$. We note that since the low order moments of $x \sim F_{yes}^{A_i,g_i,m}/Q_{j_i}$ and $x \sim F_{no}^{A_i,g_i,m}/Q_{j_i}$ are identical, these terms will cancel exactly, leaving only the Taylor error terms to contend with, which we will prove are small.

However, before we do this, we first want to deal with the outer term $2^{-B_i}\binom{B_i}{\ell}$. In particular, we show that it is at most 1. In fact,

$$\binom{B_i}{\ell}2^{-B_i} < 2^{-B_i}\sum_{\ell}\binom{B_i}{\ell}1^\ell 1^{B_i-\ell} = 1$$

by the Binomial Theorem.

We next let $f(x) := (1+x)^\ell(1-x)^{B_i-\ell}$. By Taylor expanding about $x=0$, we find that

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 \cdots + \frac{f^{(m-1)}(0)}{(m-1)!}x^{m-1} + R_m(x) \tag{1}$$

where $R_m(x) = \frac{f^m(\zeta)}{m!}x^m$ for some ζ between 0 and x.

As $E_{\delta_i \in F_{yes}}[\delta_i^k] = E_{\delta_i \in F_{no}}[\delta_i^k]$ for $k < m$ by Lemma 7, we have $E_{x \sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[f(x_i)] = E_{x \sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[f(x_i)]$ if $B_i < m$. On the other hand, if $B_i \geq m$, the expectations of the non-remainder terms over $x \sim F_{yes}^{A_i,g_i,m}/Q_{j_i}$ and $x \sim F_{no}^{A_i,g_i,m}/Q_{j_i}$ will be the same. Thus, in that case we have that

$$E_{x \sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[f(x)] - E_{x \sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[f(x)] = E_{x \sim F_{yes}^{A_i,g_i,m}/Q_{j_i}} R_m(x) - E_{x \sim F_{yes}^{A_i,g_i,m}/Q_{j_i}} R_m(x).$$

We will try to bound this by showing that $|R_m(x)|$ is small at least when ℓ is close to $B_i/2$. [[TODO: Mention bounds on ℓ.]]

Lemma 12. *Take x to be a real number with $|x| < \frac{1}{10}$, if $m \geq B_i$ and $\left|\ell - \frac{B_i}{2}\right| \leq \frac{B_i}{5}$, then*

$$|R_m(x)| \leq (1+|x|)^{|B_i-2\ell|} m^4 (2(|x|\sqrt{B_i} + |x||B_i - 2\ell| + |x|^2 B_i))^m$$

where $R_m(x)$ is given in equation (1).

Proof. By definition $R_m(x) = x^m/m! f^{(m)}(y)$ for some y between 0 and x. In particular, note that this implies $|y| < 1/10$.

Using Leibnitz Rule, the m-th derivative of $f(y)$ can be expressed as

$$\begin{aligned} f^{(m)}(y) &= \sum_{t=0}^{m}\binom{m}{t}(B_i-\ell)_{m-t}(\ell)_t(1+y)^{\ell-t}(1-y)^{B_i-\ell-m+t}(-1)^{m-t} \\ &= (1+y)^\ell(1-y)^{B_i-\ell}\left(\sum_{t=0}^{m}\binom{m}{t}(B_i-\ell)_{m-t}(\ell)_t(\frac{1}{1+y})^t(\frac{1}{1-y})^{m-t}(-1)^{m-t}\right). \end{aligned} \tag{2}$$

Note that the summand in (2) is roughly

$$\binom{m}{t}(B_i-\ell)^{m-t}\ell^t(\ell)_t(\frac{1}{1+y})^t(\frac{1}{1-y})^{m-t}(-1)^{m-t}.$$

If this were exactly true, we could use the binomial theorem to rewrite it as

$$\left(\frac{\ell}{1+y} - \frac{B_i-\ell}{1-y}\right)^m,$$

allowing us to take advantage of significant cancellation of terms. Unfortunately, the falling factorials $(B_i-\ell)_{m-t}$ and $(\ell)_t$ are not exactly equal to the relevant exponentials. However, we can make use of Stirling numbers to write them in terms of similar exponentials.

Definition 13. *The unsigned Stirling number of the first kind, usually written as $c(n,k)$ or $\left[{n \atop k}\right]$, is defined to be the number of permutations of $[n]$ with exactly k cycles. The signed Stirling number of the first kind is defined by $s(n,k) = (-1)^{n-k}c(n,k)$.*

In particular, we make use of the fact:

Fact 14. *For any non-negative integer n and real number z we have that*

$$(z)_n = \sum_k s(n,k)z^k.$$

Using this, we may rewrite (2) as

$$(1+y)^{\ell}(1-y)^{B_i-\ell}\cdot$$
$$\sum_{t=0}^{m}\binom{m}{t}\left(\sum_{h'}(B_i-\ell)^{m-t-h'}(-1)^{h'}\left[{m-t \atop m-t-h'}\right]\right)\left(\sum_{h}(\ell)^{h-s}(-1)^{h}\left[{t \atop t-h}\right]\right)\left(\frac{1}{1+y}\right)^{t}\left(\frac{-1}{1-y}\right)^{m-t}$$

Interchanging the order of summations yields

$$(1+y)^{\ell}(1-y)^{B_i-\ell}\cdot \tag{3}$$
$$\sum_{h,h'}\left(\sum_{t=0}^{m}\binom{m}{t}\left[{t \atop t-h}\right]\left[{m-t \atop m-t-h'}\right](-1)^{h+h'}(\ell)^{t-h}(B_i-\ell)^{m-t-h'}\left(\frac{1}{1+y}\right)^{t}\left(\frac{-1}{1-y}\right)^{m-t}\right).$$

To make further progress, we would like to simplify $\binom{m}{t}\left[{t \atop t-h}\right]\left[{m-t \atop m-t-h'}\right]$. Specifically, we use a lemma about Stirling numbers to find an alternative expression of $\left[{t \atop t-h}\right]$ and $\left[{m-t \atop m-t-h'}\right]$ and get cancellation of the binomial coefficients.

Lemma 15. *There exist some constants $0 \le c_{f,h} \le 1$ such that for all t,h, $\left[{t \atop t-h}\right] = \sum_{f=h+1}^{2h}(t)_f c_{f,h}$*

Proof. We analyze these Stirling numbers using a combinatorial approach. $\left[{t \atop t-h}\right]$ is the number of permutations of $[t]$ with exactly $t-h$ cycles. Such permutations should have between $h+1$ and $2h$ non fixed points. Let f be the number of non-fixed points in this permutation. Then $\left[{t \atop t-h}\right]$ can be represented as $\sum_{f=h+1}^{2h}\binom{t}{f}T_{f,f-h}$ where $T_{f,f-h}$ represents the number of permutations of f elements with no fixed points that have exactly $f-h$ cycles. Note that $T_{f,f-h} \le f!$.

We have $\left[{t \atop t-h}\right] = \sum_{f=h+1}^{2h}\frac{t!}{f!(t-f)!}T_{f,f-h} = \sum_{f=h+1}^{2h}(t)_f\frac{T_{f,f-h}}{f!}$. So taking $c_{f,h} = \frac{T_{f,f-h}}{f!}$, we are done. □

Substituting the result of Lemma 15 into Stirling numbers showing up in (3), we get that

$$\binom{m}{t}\left[{t \atop t-h}\right]\left[{m-t \atop m-t-h'}\right] = \frac{m!}{t!(m-t)!}\sum_{f=h+1}^{2h}\frac{t!}{(t-f)!}c_{f,h}\sum_{g=h'+1}^{2h'}\frac{(m-t)!}{(m-t-g)!}c_{g,h'}$$

$$= \sum_{f=h+1}^{2h}\sum_{g=h'+1}^{2h'}\frac{(m-g-f)!}{(t-f)!(m-t-g)!}(m)_{f+g}c_{f,h}c_{g,h'} = \sum_{f=h+1}^{2h}\sum_{g=h'+1}^{2h'}\binom{m-g-f}{t-f}(m)_{f+g}c_{f,h}c_{g,h'}.$$

Substituting this result into equation (3), we have $f^{(m)}(y)$ equals

$$(1+y)^{\ell}(1-y)^{B_i-\ell}\cdot$$
$$\sum_{t=0}^{m}\sum_{h,h'}\sum_{f=h+1}^{2h}\sum_{g=h'+1}^{2h'}\binom{m-g-f}{t-f}(-1)^{h+h'+g}\left(\frac{\ell}{1+y}\right)^{t-f}\left(-\frac{B_i-\ell}{1-y}\right)^{m-t-g}\frac{\ell^{f-h}(B_i-\ell)^{g-h'}}{(1+y)^f(1-y)^g}(m)_{f+g}c_{f,h}c_{g,h'}.$$

Applying the binomial theorem to the sum $\sum_{t=0}^{m}\binom{m-g-f}{t-f}\left(\frac{\ell}{1+y}\right)^{t-f}\left(-\frac{B_i-\ell}{1-y}\right)^{m-t-g}$, we can get that the above is equal to

$$(1+y)^{\ell}(1-y)^{B_i-\ell}\cdot\sum_{h,h'}\sum_{f=h+1}^{2h}\sum_{g=h'+1}^{2h'}(-1)^{h+h'+g}\left(\frac{\ell}{1+y}-\frac{B_i-\ell}{1-y}\right)^{m-g-f}\frac{\ell^{f-h}(B_i-\ell)^{g-h'}}{(1+y)^f(1-y)^g}(m)_{f+g}c_{f,h}c_{g,h'}.$$

Given $0<c_{f,h},c_{g,h'}<1$, we have that $|f^{(m)}(y)|$ is at most

$$(1+y)^{\ell}(1-y)^{B_i-\ell}\sum_{h,h'}\sum_{f=h+1}^{2h}\sum_{g=h'+1}^{2h'}\left|\frac{\ell}{1+y}-\frac{B_i-\ell}{1-y}\right|^{m-g-f}\frac{\ell^{f-h}(B_i-\ell)^{g-h'}}{(1+y)^f(1-y)^g}(m)_{f+g}.$$

Note that $(m)_{f+g}$ is at most $m!$, but vanishes if $f+g>m$. In order for this not to happen, it must be the case that $h+h'<m$. This means that there are at most m^4 non-vanishing terms in the above sum as each of h and h' can take at most m values and for each pair of values, there are at most m possibilities for each of f and g. Therefore, we have that the above is at most

$$[(1+y)^{\ell}(1-y)^{B_i-\ell}]m^4m!\max_{\substack{2h\geq f>h\geq 0\\ 2h'\geq g>h'\geq 0\\ f+g\leq m}}\left[\left|\frac{\ell}{1+y}-\frac{B_i-\ell}{1-y}\right|^{m-g-f}\frac{\ell^{f-h}(B_i-\ell)^{g-h'}}{(1+y)^f(1-y)^g}\right]. \tag{4}$$

In order to bound (4), we want to find the largest summands. For this we note that increasing f or g decreases the exponent of $\left|\frac{\ell}{1+y}-\frac{B_i-\ell}{1-y}\right|$ while increasing f increases the exponent of $\left(\frac{\ell}{1+y}\right)$ and increasing g increases the exponent of $\left(\frac{B_i-\ell}{1-y}\right)$. To make progress, we need to understand the relative sizes of these terms:

Claim. *We have that* $|\frac{\ell}{1+y}-\frac{B_i-\ell}{1-y}|\leq\frac{\ell}{1+y}$ *and* $|\frac{\ell}{1+y}-\frac{B_i-\ell}{1-y}|\leq\frac{B_i-\ell}{1-y}$.

Proof. It suffices to show that $\frac{\ell}{1+y}$ and $\frac{B_i-\ell}{1-y}$ are within a factor of two of each other. For this, we note that as $|y|<1/10$ that the ratio of $1+y$ to $1-y$ is between $9/11$ and $11/9$. Furthermore, as $|B_i-2\ell|<B_i/5$, we have that the ratio of $B_i-\ell$ to ℓ is the same as the ratio of $(\ell/B_i)+(B_i-2\ell)/B_i$ to (ℓ/B_i), which is between $4/5$ and $6/5$. Multiplying these together yields our result. □

Applying this, we find that the maximum in (4) is attained either when $f+g=m$ or when $f=2h$ and $g=2h'$. In the former case we have

$$\begin{aligned}
&\max_{\substack{2h\geq f>h\geq 0\\ 2h'\geq g>h'\geq 0\\ f+g=m}}\frac{\ell^{f-h}(B_i-\ell)^{g-h'}}{(1+y)^f(1-y)^g}\\
&=\max_{\substack{2h\geq f>h\geq 0\\ 2h'\geq g>h'\geq 0\\ f+g=m}}\left(\frac{\ell^{f-h}}{(1+y)^f}\right)\left(\frac{(B_i-\ell)^{g-h'}}{(1-y)^g}\right)\\
&\leq\max_{f+g=m}\left(\frac{\ell^{f/2}}{(1+y)^f}\right)\left(\frac{(B_i-\ell)^{g/2}}{(1-y)^g}\right)\\
&=\max\left(\frac{\sqrt{\ell}}{1+y},\frac{\sqrt{B_i-\ell}}{1-y}\right)^m.
\end{aligned}$$

In the latter case, it gives

$$\begin{aligned}
&\max_{2h+2h'\le m} \left|\frac{\ell}{1+y} - \frac{B_i-\ell}{1-y}\right|^{m-2h-2h'} \frac{\ell^h (B_i-\ell)^{h'}}{(1+y)^{2h}(1-y)^{2h'}} \\
&\le \max\left(\left|\frac{\ell}{1+y} - \frac{B_i-\ell}{1-y}\right|, \frac{\sqrt{\ell}}{1+y}, \frac{\sqrt{B_i-\ell}}{1-y}\right)^m .
\end{aligned}$$

Thus, in either case we have that $|f^{(m)}(x)|$ is at most

$$[(1+y)^{\ell}(1-y)^{B_i-\ell}] m^4 m! \max\left(\left|\frac{\ell}{1+y} - \frac{B_i-\ell}{1-y}\right|, \frac{\sqrt{\ell}}{1+y}, \frac{\sqrt{B_i-\ell}}{1-y}\right)^m .$$

To bound the maximum, we note that

$$\frac{\sqrt{\ell}}{1+y}, \frac{\sqrt{B_i-\ell}}{1-y} \le \frac{\sqrt{B_i}}{9/10} \le 2\sqrt{B_i}.$$

On the other hand,

$$\begin{aligned}
\left|\frac{\ell}{1+y} - \frac{B_i-\ell}{1-y}\right| &= \frac{1}{1-y^2}\left|(B_i-2\ell) - yB_i\right| \\
&\le 2(|B_i-2\ell| + |x|B_i).
\end{aligned}$$

Finally, note that $(1+y)(1-y) = 1-y^2 < 1$. Therefore, we have that

$$[(1+y)^{\ell}(1-y)^{B_i-\ell}] \le \max((1+y),(1-y))^{|(B_i-\ell)-\ell|} = (1+|x|)^{|B_i-2\ell|}.$$

Putting this together we have that for $|x| < 1/10$ that

$$R_m(x) = |x^m f^{(m)}(y)/m!| \le (1+|x|)^{|B_i-2\ell|} m^4 (2(|x|\sqrt{B_i} + |x||B_i-2\ell| + |x|^2 B_i))^m .$$

As desired.

□

So for any value of ℓ we have that

$$\begin{aligned}
&|\Pr(X_i^{yes} = \ell) - \Pr(X_i^{no} = \ell)| \\
&\le \left|\mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[f(x)] - \mathbb{E}_{x\sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[f(x)]\right| \\
&= \left|\mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[a_0 + a_1 x + \ldots + a_{m-1}x^{m-1} + R_m(x)] - \mathbb{E}_{x\sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[a_0 + a_1 x + \ldots + a_{m-1}x^{m-1} + R_m(x)]\right| \\
&= \left|\sum_{k=0}^{m-1}\left(\mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[a_k x^k] - \mathbb{E}_{x\sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[a_k x^k]\right) + \mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[R_m(x)] - \mathbb{E}_{x\sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[R_m(x)]\right| \\
&= \left|\mathbb{E}_{x\sim F_{yes}^{A_i,g_i,m}/Q_{j_i}}[R_m(x)] - \mathbb{E}_{x\sim F_{no}^{A_i,g_i,m}/Q_{j_i}}[R_m(x)]\right| .
\end{aligned}$$

Since for all x in either distribution, we have $|x| < 1/10$, this is at most

$$2(1+|x|)^{|B_i-2\ell|} m^4 (2(|x|\sqrt{B_i} + |x||B_i-2\ell| + |x|^2 B_i))^m$$

by Lemma 12.

While this bound is fairly good when ℓ is close to $B_i/2$, it is less useful when they are far apart. Fortunately, since $B_i \geq m > C\log(s)$ we have by a Chernoff bound that except for with probability $1/s^2$ that $\mathrm{Bin}(B_i, (1+x)/2)$ is within $O(\sqrt{B_i\log(s)})$ of $B_i(1+x)/2$. Therefore, for a sufficiently large constant A,

$$\Pr(|2X_i^{yes} - B_i| > x_{max}B_i + A\sqrt{B_i\log(s)}) < 1/s^2$$

and similarly for X_i^{no}. Therefore, we have that $d_{TV}(X_i^{yes}, X_i^{no})$ is at most

$$O(1/s^2) + \sum_{\ell:|B_i-2\ell|<x_{max}B_i+A\sqrt{B_i\log(s)}} \left|\Pr(X_i^{yes}=\ell) - \Pr(X_i^{no}=\ell)\right|.$$

Using the above to bound the differences in probabilities, we get a final bound of:

$$O(1/s^2) + m^4 O(\sqrt{B_i\log(s)} + x_{max}B_i)(1+x_{max})^{O(\sqrt{B_i\log(s)}+x_{max}B_i)} O(\sqrt{x_{max}^2 B_i\log(s)} + x_{max}^2 B_i)^m.$$

This completes our proof when $Q_{j_i} = Q_{k_i}$. In general, we can assume without loss of generality that $Q_{k_i} \geq Q_{j_i}$. We will then sub-divide the bin k_i into two sub-bins with probability masses $Q_{k_i} - Q_{j_i}$ and $Q_{j_i} - \delta_i$. We can think of taking a sample from p conditioned on lying in $\{j_i, k_i\}$ as first with probability $(Q_{k_i} - Q_{j_i})/(Q_{k_i} + Q_{j_i})$ taking a sample from the first sub-bin (and thus landing in bin k_i), and otherwise taking a sample from j_i or k_i with probabilities $(Q_{j_i} \pm \delta_i)/(2Q_{j_i})$. If we are taking B_i samples from this pair, we can imagine this as first taking $X \sim \mathrm{Bin}(B_i, (Q_{k_i} - Q_{j_i})/(Q_{k_i} + Q_{j_i}))$ samples from this extra sub-bin and then taking $B_i' = B_i - X$ samples from the remaining pair. However, we note that the distribution of samples obtained in the first bin of this pair is distributed exactly as it would have been if B_i' samples were originally taken conditioned on lying in a pair of bins with probabilities $Q_{j_i} \pm \delta$. As this situation has already been analyzed (in the case where $Q_{j_i} = Q_{k_i}$), we know that the resulting total variational distance is at most

$$O(1/s^2) + m^4 O(\sqrt{B_i'\log(s)} + x_{max}B_i')(1+x_{max})^{O(\sqrt{B_i'\log(s)}+x_{max}B_i')} O(\sqrt{x_{max}^2 B_i'\log(s)} + x_{max}^2 B_i')^m.$$

Taking the expectation of this over B_i' (which is always less than B_i) yields our full result. □

Lemma 11 provides fairly good bounds so long as B_i is not too large. However, the total number of samples N that we are taking might be substantially larger. Fortunately, we can say that with high probability that no pair of bins contains too many samples. In particular we show:

Lemma 16. *If $N > \frac{6\log s}{\min_i(Q_{j_i}+Q_{k_i})}$ and $B = 2\max_i(Q_{j_i} + Q_{k_i})N$, then if N i.i.d. samples are drawn from a distribution p taken from either D_{yes} or D_{no}, then with probability at least $1 - 1/s$ there is no pair of bins (j_i, k_i) receiving a total of more than B of these samples.*

Proof. By construction, for $D^N_{yes/no}$, $\mu_i := \mathbb{E}[B_i] = (Q_{j_i} + Q_{k_i})N > 6\log s$. Note that B_i is a sum of independent and identically distributed indicator random variables. Applying the Chernoff Bounds, $\Pr(B_i \geq (1+\delta)\mu_i) \leq e^{-\frac{\delta^2\mu_i}{3}}$. Letting $\delta = 1$, then we have that $\Pr(B_i \geq 2\mu_i) \leq \frac{1}{s^2}$. As $B \geq 2\mu_i$, this says that with probability at least $1 - 1/s^2$ that the ith pair of bins does not contain more than B samples. Our result now follows by taking a union bound over i. □

2.3 Proof of Proposition 9

We are now ready to prove the full Proposition 9.

Proof. Let B_i be the number of samples from the ith pair of bins, j_i and k_i. Define U to be the vector of values $(B_1, \cdots, B_s)$ as well as the number of samples in each unpaired bin. By our construction, the distribution of U is the same for any p taken from D_{yes} or D_{no}, independently of δ_is, as $p_{j_i} + p_{k_i}$ is always $Q_{j_i} + Q_{k_i}$. So

$$d_{TV}(D_{yes}^N, D_{no}^N) = E_U[d_{TV}(D_{yes}^N|U, D_{no}^N|U)] \leq \Pr_U(\exists i : B_i \geq B) + \max_{U:B_i<B \text{ for all } i} d_{TV}(D_{yes}^N|U, D_{no}^N|U). \quad (5)$$

For the first term, we note that Lemma 16 implies that the probability that some B_i is more than B is at most $1/s$. To deal with the second term, we note that after conditioning on U, either D_{yes}^N or D_{no}^N, we observe that the choice of δ_is are independent for each pair of bins. This implied that conditioned on U, the number of samples drawn from each of j_i and k_i are independent for each i. As the distributions D_{yes}^N and D_{no}^N are symmetric in the sense that seeing a collection of samples in some order is as likely as seeing those samples in any other order, the total variational distance between these distributions conditioned on U is the same as the variational distance between their distributions over the counts of numbers of samples from each bin. These distributions in turn are product distributions over pairs of bins, and thus we can bound (5) by

$$\max_{U:B_i\leq B} \sum_{i=1}^{s} d_{TV}(X_{i,yes}^{B_i}, X_{i,no}^{B_i}) + \frac{1}{s}, \quad (6)$$

where $X_{i,yes}^{B_i}$ is the distribution over the number of samples a random p from D_{yes} draws from j_i conditioned on the fact that it drew a total of B_i samples from $\{j_i, k_i\}$, and $X_{i,no}^{B_i}$ is defined similarly. However, by Lemma 11 and the fact that $B_i \leq B$, the ith term in this sum is at most

$$O(1/s^2) + m^4 O(\sqrt{B\log(s)} + x_{max}B)(1 + x_{max})^{O(\sqrt{B\log(s)}+x_{max}B)} O(\sqrt{x_{max}^2 B\log(s)} + x_{max}^2 B)^m.$$

Summing over all i from 1 to s and adding in the extra $1/s$ term gives our final bound of

$$O(1/s) + m^4 s O(\sqrt{B\log(s)} + x_{max}B)(1 + x_{max})^{O(\sqrt{B\log(s)}+x_{max}B)} O(\sqrt{x_{max}^2 B\log(s)} + x_{max}^2 B)^m.$$

□

3 One-Dimensional Monotonicity Testing

As a warmup we will prove the $d = 1$ version of Theorem 4.

3.1 Construction

In this section, we focus on getting a new lower bound of testing monotone distribution over $[n]$. We will do this by producing a version of the construction in Section 2 so that a distribution from D_{yes} is always monotone and a distribution from D_{no} is far from monotone with high probability. We begin by proving it for n not too large.

In particular, assume that n is an even number with $C^2\log(1/\epsilon) < n < \frac{1}{C^4(\log\frac{1}{\epsilon})^3\epsilon}$ for some sufficiently large constant C and assume that ϵ is sufficiently small. We begin with defining a base distribution Q over $S = [n]$ by

$$Q_{2i-1} = Q_{2i} = \frac{5}{4n} + \frac{1}{2n^2} - \frac{i}{n^2}, 1 \leq i \leq \frac{n}{2}.$$

Note that $Q_i \geq Q_j$ for any $i \leq j$ and $\sum_{i=1}^{n} Q_i = 1$. We define the sequence of pairs $(j_i, k_i)_{1\leq i\leq \frac{n}{2}}$ by $j_i = 2i-1, k_i = 2i$. Note that these cover all bins in S exactly once.

To complete the construction, we need to define values of A_i, g_i, and m. In particular, we let m be the smallest odd integer that is more than $C\log(1/\epsilon)$.

Note that we have $nm^3 < 1/(C\epsilon)$, and thus $\frac{1}{4n^2} > \frac{8m^3\epsilon}{n}$. Therefore, taking $A_i = A = \frac{8m^3\epsilon}{n}$ for all i, we have $\frac{1}{4n^2} > A$. We also let $g_i = \cos(\frac{\pi}{m})$ for all i. Since A_i and g_i are both constants, the distributions of $F^i_{yes/no}$ are identical for all i. For convenience of notation, we denote this distribution to be $F_{yes/no}$. Given $\epsilon > 0$ sufficiently small, we can assume that n and m are larger than sufficiently large constants. It's easy to check that this construction satisfies $A < \frac{\min_i Q_i}{2+2|g|}$.

In order to prove the monotonicity/non-monotonicity of D_{yes}/D_{no} we will need some properties of the $F_{yes/no}$ with these particular parameters. In fact, we will prove a slightly more general form:

Lemma 17. *For m a sufficiently large positive odd integer, $g = \cos(\frac{\pi}{m})$, and $A > 0$, $F^{A,g,m}_{yes}$ and $F^{A,g,m}_{no}$ have the following properties:*

1. *If δ is taken from either distribution $|\delta| < 2A$ almost surely.*
2. *If δ is taken from $F^{A,g,m}_{yes}$, then $\delta \geq 0$ almost surely.*
3. *If δ is taken from $F^{A,g,m}_{no}$, then there is a probability of $1/m$ that δ is negative, in which case we have $\delta < -A/m^2$.*

Proof. Property 1 follows from the fact that the cosine terms all have absolute value at most 1. For property 2, since m is odd, we have that $\cos(\frac{2\pi a}{m}) \geq \cos(\frac{(m-1)\pi}{m}) = -\cos(\frac{\pi}{m})$, so all δs drawn from F_{yes} will be non-negative. For distribution of F_{no}, we have $A(\cos(\frac{2\pi(a+\frac{1}{2})}{m}) + \cos(\frac{\pi}{m})) < 0$ if and only if $a = \frac{m-1}{2}$. Since there are m choices of a, we conclude that δ drawn from F_{no} will have $\frac{1}{m}$ chance to be negative, and the negative value is $A(\cos(\frac{\pi}{m}) - 1)$. By Taylor expanding $\cos(x)$ about 0, we find that it is at most $-A(\frac{\pi^2}{2m^2} - \frac{\pi^4}{24m^4})$. Given that m is sufficiently large this is at most $-A/m^2$. □

Lemma 17 ensures that δs drawn from either distribution are small. In addition, it guarantees that δs drawn from F_{yes} are positive with probability 1 and δs drawn from F_{no} are negative with probability $\frac{1}{m}$. This makes sure that the distributions in D_{yes} and the distributions in D_{no} are different in terms of being monotone or not, as we can see in the later analysis.

We want to show that a random distribution drawn from D_{yes} is monotone, and a random distribution drawn from D_{no} is some distance from monotone with high probability. This will mean that any monotonicity tester will be able to distinguish between a distribution from D_{yes} and a distribution from D_{no}, which is impossible without many samples by Proposition 9.

Lemma 18. *A distribution p drawn from D_{yes} is monotone with probability 1.*

Proof. We will prove that $p_i \geq p_{i+1}$ for all i. Firstly, we note that for odd i, $Q_i = Q_{i+1}$. According to construction, $p_i = Q_i + \delta_{\frac{i+1}{2}}$ and $p_{i+1} = Q_i - \delta_{\frac{i+1}{2}}$. Applying Lemma 17, for $\delta \in F_{yes}$, $\delta \geq 0$ gives $p_i \geq p_{i+1}$.

For even i, $p_i = \frac{5}{4n} + \frac{1}{2n^2} - \frac{i}{2n^2} - \delta_{\frac{i}{2}}$ and $p_{i+1} = \frac{5}{4n} + \frac{1}{2n^2} - \frac{i+2}{2n^2} + \delta_{\frac{i+2}{2}}$. Thus,

$$p_i - p_{i+1} = \frac{1}{n^2} - (\delta_{\frac{i}{2}} + \delta_{\frac{i+2}{2}})$$

By Lemma 17, $|\delta_j| < 2A < 1/(2n^2)$ for each j. Thus

$$p_i - p_{i+1} \geq \frac{1}{n^2} - \frac{2}{2n^2} = 0.$$

This completes our proof. □

In contrast to distributions in D_{yes}, the distributions in D_{no} is at least ϵ far from monotone with high probability. In order to show this, we first need a lemma allowing us to show that some distributions p are far from *any* monotone distribution.

Lemma 19. *For a distribution p, and q an arbitrary monotone distribution, then $d_{TV}(p,q) \geq \sum_{i=1}^{\frac{n}{2}} \gamma_i$, where*

$$\gamma_i = \begin{cases} |p_{2i-1} - p_{2i}|/2 & \textit{if } p_{2i-1} - p_{2i} < 0 \\ 0 & \textit{if } p_{2i-1} - p_{2i} \geq 0. \end{cases}$$

Proof. We will show that $(|q_{2i-1} - p_{2i-1}| + |q_{2i} - p_{2i}|)/2 \geq \gamma_i$. In particular, if $p_{2i-1} - p_{2i} \geq 0$, $\gamma_i = 0$ and we have our desired inequality. If $p_{2i-1} - p_{2i} < 0$, then $\gamma_i = (p_{2i} - p_{2i-1})/2 > 0$. By definition, q monotone implies $q_{2i-1} \geq q_{2i}$. This means that

$$|q_{2i-1} - p_{2i-1}| + |q_{2i} - p_{2i}| \geq (q_{2i-1} - p_{2i-1}) + (p_{2i} - q_{2i}) = q_{2i-1} - q_{2i} + 2\gamma_i \geq 2\gamma_i.$$

Summing this inequality over all i, we have:

$$d_{TV}(p,q) = \frac{1}{2}\sum_{i=1}^{n} |q_i - p_i| = \sum_{i=1}^{\frac{n}{2}} ((|q_{2i-1} - p_{2i-1}| + |q_{2i} - p_{2i}|))/2 \geq \sum_{i=1}^{\frac{n}{2}} \gamma_i.$$

This completes our proof. □

We can now use Lemma 19 to show that a distribution from D_{no} is far from monotone with high probability.

Lemma 20. *With 99% probability, a random distribution drawn from D_{no} is ϵ far from monotone.*

Proof. Let γ_i be as in Lemma 19, we have that

$$\gamma_i = \begin{cases} |p_{2i-1} - p_{2i}|/2 & \text{if } p_{2i-1} - p_{2i} < 0 \\ 0 & \text{if } p_{2i-1} - p_{2i} \geq 0. \end{cases}$$

Note that $p_{2i-1} - p_{2i} = (Q_{2i-1} + \delta_i) - (Q_{2i} - \delta_i) = 2\delta_i$. Thus we have that

$$\gamma_i = \begin{cases} -\delta_i & \text{if } \delta_i < 0 \\ 0 & \text{if } \delta_i \geq 0. \end{cases}$$

By Lemma 17, we have that each γ_i is positive (with absolute value at least $8\epsilon m/n$) independently with probability $1/m$. Let X be the number of these positive terms. We have that $X \sim \text{Bin}(n/2, 1/m)$. As n/m is at least a large constant, we have with 99% probability that $X > n/(4m)$. If this holds then by Lemma 19 the distance of p from the nearest monotone distribution is at least

$$\sum_{i=1}^{n/2} \gamma_i \geq (n/(4m))(8\epsilon m/n) > \epsilon.$$

This completes our proof. □

We are now prepared to prove Theorem 4 when $d = 1$ and $n < \frac{1}{C^4(\log\frac{1}{\epsilon})^3\epsilon}$ is even. Let N be a sufficiently small multiple of $n/(m^6\log(n)\epsilon^2)$ and suppose for sake of contradiction that there is a monotonicity algorithm with a probability 2/3 of success using only N samples. Running this algorithm should be able to distinguish N samples taken from a random distribution from D_{yes} and a random distribution from D_{no} with probability of success at least 3/5 since the distribution in the former case will be monotone, and the distribution in the latter will be at least ϵ-far from monotone with probability at least 99%.

On the other hand, we can apply Corollary 10 here as $B = \Theta(N/n)$ and $x_{max} = O(An) = O(m^3\epsilon)$. Thus, $Bx_{max}^2 = O(Nm^3\epsilon^2/n)$, which is at most a small multiple of $1/\log(s)$. This implies that

$d_{TV}(D^N_{yes}, D^N_{no}) < 1/100$, and thus the difference in the probability that our tester accepts a distribution from D_{yes} given N samples can differ from the probability of accepting a distribution from D_{no} given N samples by at most $1/100$.

This completes the proof when n is even and at most $\frac{1}{C^4(\log\frac{1}{\epsilon})^3\epsilon}$. For other n, we let n_0 be the largest even number less than both n and $\frac{1}{C^4(\log\frac{1}{\epsilon})^3\epsilon}$. We note that a monotonicity tester on $[n]$ can be used to obtain a monotonicity tester on $[n_0]$ simply by ignoring the extra bins in the domain. Thus, we get a lower bound of $\Omega(n_0/(\log^7(1/\epsilon)\epsilon^2)) = \Omega(\min(n, (1/\epsilon)/\log^3(1/\epsilon))/(\epsilon\log^7(1/\epsilon)))$.

This completes our proof.

4 Multidimensional Monotonicity Testing

In this section, we generalize the results of the previous section to cover d-dimensional monotonicity testing.

4.1 Construction

For the one dimensional case we were able to modify our monotone base distribution Q to make it non-monotone by exchanging bits of probability mass between adjacent bins. In the high dimensional case however, it is not clear what the appropriate generalization of this should be, especially given that there are pairs of bins $\mathbf{i}$ and $\mathbf{j}$ that are incomparable to each other in the relevant ordering. Thus, in order to construct $D_{yes/no}$ for the multidimensional case, we have to find comparable pairs of bins to move. Here we introduce the notion of cubes and halfcubes for a distribution over $[n]^d$ so that most bins in a halfcube are comparable to a bin in another halfcube within the same cube. Once again, we start by proving it when n is not too large.

Suppose $Cd\log(1/\epsilon) < n < \frac{d}{(C^2\log\frac{1}{\epsilon})^3\epsilon}$, and $d < (C^2\log\frac{1}{\epsilon})^3$ for some sufficiently large constant C and that $2d|n$. We begin by separating a distribution over $[n]^d$ into $(\frac{n}{2d})^d$ cubes with each of them having $(2d)^d$ bins, with the idea of using these cubes as a unit to replace the pairs of bins in the one dimensional construction. More formally,

Definition 21. *Let $\boldsymbol{1}$ be the d-dimensional vector $(1, 1, 1..., 1)$. For $\boldsymbol{i}, \boldsymbol{j} \in [n]^d$, we denote $\boldsymbol{i} < \boldsymbol{j}$ when $\boldsymbol{i}_a < \boldsymbol{j}_a$ for all $1 \le a \le d$.*

For a distribution over $[n]^d$, we define its $\mathbf{i}$th cube (for some $\boldsymbol{i} \in [\frac{n}{2d}]^d$) to be the set of $\mathbf{j}$th bins where $2d(\boldsymbol{i} - \boldsymbol{1}) < \boldsymbol{j} < 2d\boldsymbol{i} + \boldsymbol{1}$. Within the $\mathbf{i}$th cube, we define $J_{\boldsymbol{i}}$ to be the set of bins $\{\boldsymbol{j} : \boldsymbol{j}_1 \le 2d\boldsymbol{i}_1 - d\}$ and $K_{\boldsymbol{i}} = \{\boldsymbol{j} : \boldsymbol{j}_1 > 2d\boldsymbol{i}_1 - d\}$. We call $J_{\boldsymbol{i}}$ the first halfcube of the $\mathbf{i}$th cube and $K_{\boldsymbol{i}}$ the second halfcube.

Note that we are separating the $\mathbf{i}$th cube into 2 halfcubes based on the magnitude of its first coordinate, so $|J_{\mathbf{i}}| = |K_{\mathbf{i}}|$ and the $\mathbf{i}$th cube is $J_{\mathbf{i}} \cup K_{\mathbf{i}}$. Our construction will produce distributions that are uniform over each halfcube, so we can construct our base distribution over these halfcubes. In particular, we will use an instantiation of the ensembles from Section 2 to produce these distributions over halfcubes.

Let the distribution Q over $[\frac{n}{2d}]^d \times \{F, S\}$ given by

$$Q_{\mathbf{i},F} = Q_{\mathbf{i},S} := \frac{5(2d)^d}{8n^d} + \frac{2^d d^{d+1}}{4n^{d+1}} - \frac{(\mathbf{i}_1 + \mathbf{i}_2 + ...\mathbf{i}_d)2^{d-1}d^d}{n^{d+1}}, \mathbf{i} = (\mathbf{i}_1, \mathbf{i}_2, ...\mathbf{i}_d) \in [\frac{n}{2d}]^d.$$

Summing over all $\mathbf{i}_1$s and multiply by d, we have

$$\sum_{\mathbf{i}\in[\frac{n}{2d}]^d} (\mathbf{i}_1 + \mathbf{i}_2 + ...\mathbf{i}_d) = (1 + \frac{n}{2d})\frac{n}{4d}(\frac{n}{2d})^{d-1}d = (\frac{d}{2} + \frac{n}{4})(\frac{n}{2d})^d.$$

Therefore, we get

$$\sum_{\mathbf{i}\in[\frac{n}{2d}]^d} Q_{\mathbf{i},F} + \sum_{\mathbf{i}\in[\frac{n}{2d}]^d} Q_{\mathbf{i},S} = 2(\frac{5}{8} + \frac{d}{4n} - (\frac{d}{2} + \frac{n}{4})(\frac{n}{2d})^d \frac{2^{d-1}d^d}{n^{d+1}}) = 1.$$

This and the fact that

$$Q_{\mathbf{i},F} = Q_{\mathbf{i},S} \geq \frac{5(2d)^d}{8n^d} - \frac{d(n/2d)2^{d-1}d^d}{n^{d+1}} > (1/4)(2d/n)^d$$

shows that Q is a valid probability distribution.

As in Definition 3, we define the sequence $(j_{\mathbf{i}}, k_{\mathbf{i}})_{\mathbf{i}\in[\frac{n}{2d}]^d}$ where $j_{\mathbf{i}} = (\mathbf{i}, F)$ and $k_{\mathbf{i}} = (\mathbf{i}, S)$ to specify which halfcube of bins we are moving. Then we construct $F^{\mathbf{i}}_{yes}$ and $F^{\mathbf{i}}_{no}$ by choosing proper $A_{\mathbf{i}}, g_{\mathbf{i}}, m$:

In particular, let m be the smallest odd integer that is larger than $C\log(1/\epsilon)$ where C is a sufficiently large constant. The fact that $n < \frac{d}{(C^2\log\frac{1}{\epsilon})^3\epsilon}$ and $d < (C^2\log\frac{1}{\epsilon})^3$ imply that $nm^3 < \frac{d}{C\epsilon}$, and $\frac{2^{d-3}d^{d+1}}{n^{d+1}} > \frac{2^{d+2}m^3d^d\epsilon}{n^d}$. We take $A = \frac{2^{d+2}m^3d^d\epsilon}{n^d}$, noting that $\frac{2^{d-3}d^{d+1}}{n^{d+1}} > A$, and let $A_{\mathbf{i}} = A$ for all $\mathbf{i}$. Assuming $\epsilon > 0$ is sufficiently small (as otherwise there is nothing to prove), we may assume that n and m are at least sufficiently large constants. Let $g_{\mathbf{i}} = \cos(\frac{\pi}{m})$ for all $\mathbf{i}$.

As the $A_{\mathbf{i}}$ and $g_{\mathbf{i}}$ are the same for all $\mathbf{i}$, we refer to $F^{A_{\mathbf{i}},g_{\mathbf{i}},m}_{yes/no}$ simply as $F_{yes/no}$ and we note that Lemma 17 applies to them. Noting that $A < \min(Q_{\mathbf{i},F}, Q_{\mathbf{i},S})/(1+|g|)$, we can invoke Definition 8 to define ensembles C_{yes} and C_{no} over the set of halfcubes. Using these we construct our actual hard instances over $[n]^d$ as follows:

Definition 22. *We define ensembles $D_{yes/no}$ of distributions over $[n]^d$ in the following way: To sample a distribution q from $D_{yes/no}$: first get a sample distribution p over halfcubes from $C_{yes/no}$, one then takes a sample from q by first sampling a halfcube using p and then returning a uniform random sample from that halfcube.*

Note that in this construction, any distribution $q \sim D_{yes/no}$ is uniform inside each halfcube. And if we consider a distribution p over $[\frac{n}{2d}]^d \times \{F, S\}$ where $p_{\mathbf{i},F} = \frac{(2d)^d}{2} q_{\mathbf{j}}$ where $\mathbf{j} \in J_{\mathbf{i}}$ and $p_{\mathbf{i},S} = \frac{(2d)^d}{2} q_{\mathbf{j}}$ where $\mathbf{j} \in K_{\mathbf{i}}$, we have $p \in C_{yes/no}$ by definition. Since one can produce a sample from q given a sample from p, the statistical task of distinguishing whether a distribution q was taken from D_{yes} or D_{no} in N samples is equivalent to the task of distinguishing whether a distribution p was taken from C_{yes} or C_{no} in N samples. We will show by Corollary 10 that this latter task is hard unless N is large, but first we need to prove that a distribution in D_{yes} is monotone with probability 1.

Lemma 23. *A distribution q drawn from D_{yes} is monotone with probability 1.*

Proof. Firstly, we note that since we are separating halfcubes based on the magnitude of its first coordinate, for any pair of bins $\mathbf{j}$ and $\mathbf{k}$ in the $\mathbf{i}$th cube where $\mathbf{j} < \mathbf{k}$, $\mathbf{j}$ and $\mathbf{k}$ are either in the same halfcube or $\mathbf{j}$ is in the first halfcube and $\mathbf{k}$ in the second halfcube. If they are in the same halfcube, $q_{\mathbf{j}} = q_{\mathbf{k}}$ by construction. For the case that $\mathbf{j}$ is in the first halfcube and $\mathbf{k}$ is in the second halfcube, we show that $q_{\mathbf{j}} \geq q_{\mathbf{k}}$ by proving that within each cube, any bin in the first halfcube is always heavier than any bin in the second halfcube. Note that a bin in the first halfcube in the $\mathbf{i}$th cube has weight $\frac{Q_{\mathbf{i},F}+\delta_{\mathbf{i}}}{2^{d-1}d^d}$ and the one in the second halfcube in the $\mathbf{i}$th pair has weight $\frac{Q_{\mathbf{i},S}-\delta_{\mathbf{i}}}{2^{d-1}d^d}$. Since $\delta_{\mathbf{i}} \geq 0$ for $\delta_{\mathbf{i}}$ drawn from F_{yes} (by Lemma 17) and $Q_{\mathbf{i},S} = Q_{\mathbf{i},F}$ for all $\mathbf{i}$, $\frac{Q_{\mathbf{i},F}+\delta_{\mathbf{i}}}{2^{d-1}d^d} \geq \frac{Q_{\mathbf{i},S}-\delta_{\mathbf{i}}}{2^{d-1}d^d}$ always holds. So for any pair of bins $\mathbf{j}$ and $\mathbf{k}$ in the $\mathbf{i}$th cube where $\mathbf{j} < \mathbf{k}$, $q_{\mathbf{j}} \geq q_{\mathbf{k}}$.

Secondly, we want to make sure the distribution is monotone across cubes. Note that there exists bins in the $\mathbf{i}$th cube that are comparable to some bins in the $\mathbf{j}$th cube for $\mathbf{j} \neq \mathbf{i}$ if and only if $\mathbf{i}_a \leq \mathbf{j}_a$ for $1 \leq a \leq d$ and there exists a such that $\mathbf{i}_a < \mathbf{j}_a$. As the previous paragraph implies that bins in the first half of each cube are heavier than bins in the second half, it suffices to prove that any bin in the second halfcube of the

$\mathbf{i}$th cube (with weight $\frac{Q_{\mathbf{i},S}-\delta_{\mathbf{i}}}{2^{d-1}d^d}$) is heavier than any bin in the first halfcube in the $\mathbf{j}$th cube (with weight $\frac{Q_{\mathbf{j},F}+\delta_{\mathbf{j}}}{2^{d-1}d^d}$).

We note that the difference,

$$\frac{Q_{\mathbf{i},S}-\delta_{\mathbf{i}}}{2^{d-1}d^d} - \frac{Q_{\mathbf{j},F}+\delta_{\mathbf{j}}}{2^{d-1}d^d}$$

equals

$$\frac{1}{2^{d-1}d^d}\left[\left(\frac{5(2d)^d}{8n^d}+\frac{2^d d^{d+1}}{4n^{d+1}}-\frac{(\mathbf{i}_1+\mathbf{i}_2+...\mathbf{i}_d)2^{d-1}d^d}{n^{d+1}}\right)-\left(\frac{5(2d)^d}{8n^d}+\frac{2^d d^{d+1}}{4n^{d+1}}-\frac{(\mathbf{j}_1+\mathbf{j}_2+...\mathbf{j}_d)2^{d-1}d^d}{n^{d+1}}\right)-\delta_{\mathbf{i}}-\delta_{\mathbf{j}}\right]$$

$$\geq \frac{1}{2^{d-1}d^d}\left(\frac{2^{d-1}d^{d+1}}{n^{d+1}}-\delta_{\mathbf{i}}-\delta_{\mathbf{j}}\right)$$

By Lemma 17, we have $\delta_{\mathbf{i}} < \frac{2^{d-2}d^{d+1}}{n^{d+1}}$ for $\delta_{\mathbf{i}} \sim F_{yes}$, so $\delta_{\mathbf{i}}+\delta_{\mathbf{j}} < \frac{2^{d-1}d^{d+1}}{n^{d+1}}$, which completes the proof that all distributions in D_{yes} are monotone. □

In contrast to distributions in D_{yes}, the distributions in D_{no} is at least ϵ far from monotone with high probability. To show this, we first need to prove a lemma about how far a distribution which is uniform over halfcubes is from monotone.

Lemma 24. *For a distribution q uniform within each halfcube and p an arbitrary monotone distribution, $d_{TV}(p,q) \geq \frac{1}{2}\sum_{\mathbf{i}\in[\frac{n}{2d}]^d}\gamma_i$, where $\gamma_i = (\frac{2d-1}{2d})^{d-1}\max(0, \sum_{j\in K_i} q_j - \sum_{j\in J_i} q_j)$*

Proof. Define $S_{\mathbf{i},1} = \{\mathbf{j} : \mathbf{j} \in J_{\mathbf{i}} \text{ and } \mathbf{j}_a \neq 2d\mathbf{i}_a, 2 \leq a \leq d\}$ and $S_{\mathbf{i},2} = \{\mathbf{j} : \mathbf{j} \in K_{\mathbf{i}} \text{ and } j_a \neq 2d\mathbf{i}_a - 2d + 1, 2 \leq a \leq d\}$, we can pair a bin $\mathbf{j} \in S_{\mathbf{i},1}$ to $\mathbf{k} \in S_{\mathbf{i},2}$ where $\mathbf{j} + (d, 1, 1, \ldots, 1) = \mathbf{k}$, so that for each such pair $\mathbf{j} < \mathbf{k}$. The probability that a random bin in the first halfcube of the $\mathbf{i}$th cube lies in $S_{\mathbf{i},1}$ is $(\frac{2d-1}{2d})^{d-1}$. Similarly, the probability that a random bin in the second halfcube of the $\mathbf{i}$th cube lies in $S_{\mathbf{i},2}$ is $(\frac{2d-1}{2d})^{d-1}$. Given that q is uniform within each halfcube, we can get $\gamma_{\mathbf{i}} = \max(0, \sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}} - \sum\limits_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}})$.

Note that since $S_{\mathbf{i},1} \cup S_{\mathbf{i},2} \subset J_{\mathbf{i}} \cup K_{\mathbf{i}}$,

$$d_{TV}(p,q) = \frac{1}{2}\sum_{\mathbf{i}\in[\frac{n}{2d}]^d}\sum_{\mathbf{j}\in J_{\mathbf{i}}\cup K_{\mathbf{i}}}|p_{\mathbf{j}}-q_{\mathbf{j}}| \geq \frac{1}{2}\sum_{\mathbf{i}\in[\frac{n}{2d}]^d}\sum_{\mathbf{j}\in S_{\mathbf{i},1}\cup S_{\mathbf{i},2}}|p_{\mathbf{j}}-q_{\mathbf{j}}| \geq \frac{1}{2}\sum_{\mathbf{i}\in[\frac{n}{2d}]^d}\left(\left|\sum_{\mathbf{j}\in S_{\mathbf{i},1}} p_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}}\right| + \left|\sum_{\mathbf{j}\in S_{\mathbf{i},2}} p_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}}\right|\right)$$

It suffices to prove that $\left(\left|\sum\limits_{\mathbf{j}\in S_{\mathbf{i},1}} p_{\mathbf{j}} - \sum\limits_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}}\right| + \left|\sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} p_{\mathbf{j}} - \sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}}\right|\right) \geq \gamma_{\mathbf{i}}$ for all $\mathbf{i}$. Given that p is monotone and $\mathbf{k} > \mathbf{j}$, $p_{\mathbf{j}} \geq p_{\mathbf{k}}$ for all pairs of $\mathbf{j}$ and $\mathbf{k}$ in $S_{\mathbf{i},1}$ and $S_{\mathbf{i},2}$ with $\mathbf{k} = \mathbf{j} + (d, 1, \ldots, 1)$. Therefore, summing over $S_{\mathbf{i},1}$ and $S_{\mathbf{i},2}$, $\sum\limits_{\mathbf{j}\in S_{\mathbf{i},1}} p_{\mathbf{j}} \geq \sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} p_{\mathbf{j}}$ for all $\mathbf{i}$.

If $\sum_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}} - \sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}} \geq 0$, then $\gamma_{\mathbf{i}} = 0$, and we have our desire inequality.

If $\sum_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}} - \sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}} < 0$, then $\gamma_{\mathbf{i}} = \sum\limits_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}} - \sum\limits_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}} > 0$. In this case,

$$\begin{aligned}|\sum_{\mathbf{j}\in S_{\mathbf{i},1}} p_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}}| + |\sum_{\mathbf{j}\in S_{\mathbf{i},2}} p_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}}| &\geq (\sum_{\mathbf{j}\in S_{\mathbf{i},1}} p_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}}) + (\sum_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},2}} p_{\mathbf{j}}) \\ &= (\sum_{\mathbf{j}\in S_{\mathbf{i},1}} p_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},2}} p_{\mathbf{j}}) + (\sum_{\mathbf{j}\in S_{\mathbf{i},2}} q_{\mathbf{j}} - \sum_{\mathbf{j}\in S_{\mathbf{i},1}} q_{\mathbf{j}}) \geq \gamma_{\mathbf{i}}.\end{aligned}$$

Summing this inequality over all $\mathbf{i}$, we have that $d_{TV}(p,q) \geq \frac{1}{2}\sum_{\mathbf{i}\in[\frac{n}{2d}]^d}\gamma_{\mathbf{i}}$. □

This lemma is a multidimensional analogue of Lemma 19. It gives a lower bound of distance from monotone for a distribution uniform within each halfcube, which helps us prove that a random distribution from D_{no} is not close to being monotone. We will prove in the next lemma that it's at least ϵ far from monotone with probability at least 99%.

Lemma 25. *With 99% probability, a random distribution drawn from D_{no} is at least ϵ far from monotone.*

Proof. By the definition of D_{no}, if q is sampled from D_{no}, we have that

$$\begin{aligned}\sum_{\mathbf{j}\in K_{\mathbf{i}}} q_{\mathbf{j}} - \sum_{\mathbf{j}\in J_{\mathbf{i}}} q_{\mathbf{j}} = p_{\mathbf{i},S} - p_{\mathbf{i},F} &= \\ &= (Q_{\mathbf{i},S} - \delta_{\mathbf{i}}) - (Q_{\mathbf{i},F} + \delta_{\mathbf{i}}) \\ &= -2\delta_{\mathbf{i}}.\end{aligned}$$

Where p is the corresponding distribution from C_{no}. Therefore, in the notation of Lemma 24 we have that

$$\gamma_{\mathbf{i}} = \begin{cases} 0 & \text{if } \delta_{\mathbf{i}} \geq 0 \\ -2\left(\frac{2d-1}{2d}\right)^{d-1}\delta_{\mathbf{i}} & \text{if } \delta_{\mathbf{i}} < 0. \end{cases}$$

Note that $\left(\frac{2d-1}{2d}\right)^{d-1} = \frac{1}{(1+1/(2d-1))^{d-1}} > e^{-1/2}$. Thus, by Lemma 17, this means that $\gamma_{\mathbf{i}}$ is non-zero independently with probability $1/m$ and if it is non-zero, it is at least $2^{d+2}md^d\epsilon/n^d$.

Letting X be the number of non-zero $\gamma_{\mathbf{i}}$'s. It is distributed as $\mathrm{Bin}((n/2d)^d, 1/m)$ and so with probability at least 99% is at least $(n/2d)^d/(2m)$. In such a case we have that the distance of q from uniform is at least

$$\frac{1}{2}\sum_{\mathbf{i}\in[\frac{n}{2d}]^d} \gamma_{\mathbf{i}} \geq X2^{d+1}md^d\epsilon/n^d > \epsilon.$$

□

We have proved that a distribution in D_{yes} is monotone with probability 1 and a distribution in D_{no} is ϵ far from monotone with 99% probability. In the next section, we will apply Proposition 9 to use this to show that one cannot build a monotonicity tester with too few samples.

4.2 Lower Bound of Multidimensional Monotonicity Testing

Here we prove Theorem 4 starting with the case where n is at most $\frac{d}{(C^2\log\frac{1}{\epsilon})^3\epsilon}$ and is a multiple of $2d$. Let N be a sufficiently small multiple of $(n/2d)^d(1/\epsilon)^2/(dm^6\log(1/\epsilon))$ and suppose for sake of contradiction that there is a tester that tests monotonicity over $[n]^d$ with N samples. As a distribution from D_{yes} is monotone and a distribution from D_{no} is ϵ-far from monotone with 99% probability, this tester can reliably distinguish N samples from a distribution from D_{yes} from N samples from a distribution from D_{no}. However, this is equivalent to distinguishing C_{yes} from C_{no}, which is difficult by Proposition 9.

In particular, in the context of Corollary 10, we have that $B = O(N(n/2d)^d)$ and $x_{max} = O(m^3\epsilon) < 1/10$. This makes Bx_{max}^2 at most a small multiple of $1/(d\log(1/\epsilon)) < \log(s)$. Thus, we can apply Corollary 10 and conclude that $d_{TV}(C_{yes}^N, C_{no}^N) < 1/100$ and thus that one cannot distinguish the two with N samples, providing our contradiction.

For n not of the desired form, let n_0 be the largest integer smaller than n that is both a multiple of $2d$ and at most $\frac{d}{(C^2\log\frac{1}{\epsilon})^3\epsilon}$. As monotonicity testing over $[n_0]^d$ is a special case of monotonicity testing over $[n]^d$, we obtain a lower bound of

$$\Omega((n_0/2d)^d(1/\epsilon)^2/(dm^6\log(1/\epsilon))) = 2^{-O(d)}d^{-d}\epsilon^{-2}\log^{-7}(1/\epsilon)\min(n, d\epsilon^{-1}\log^{-3}(1/\epsilon))^d.$$

This completes our proof.

5 Log Concavity Distribution Testing

To prove our lower bound for log-concavity testing, we will again use Proposition 9 to construct indistinguishable ensembles D_{yes} and D_{no}. In particular, we need to carefully instantiate our construction so that distributions from D_{yes} are log-concave almost surely, while distributions from D_{no} are ϵ far with high probability. Then Proposition 1 will imply that these ensembles are indistinguishable without a large number of samples, giving us a desired lower bound.

5.1 Construction

The intuition for log concavity testing over $[n]$ is: start with a log concave base distribution Q over $[n]$ and separate it into groups of 6 bins, then modify the 2nd and 5th bin in each group. The reason we choose to move bins in this way is that with only 1 bin being moved in each triple, it's easier to evaluate how each move affects log concavity. As long as the size of the move is small enough, the distribution will still be log concave. On the other hand, large moves cause it to be ϵ far from log concavity.

We begin by constructing the base distribution Q over $[n]$. Firstly, we assume that with $C\log(1/\epsilon) < n < \frac{1}{C^2\epsilon^{\frac{1}{2}}(\log\frac{1}{\epsilon})^{\frac{3}{2}}}$ for some sufficiently large constant C and that n is a multiple of 6. We let $Q_i = \frac{b}{n}e^{-(\frac{i}{n})^2}$ with $b = \frac{n}{\sum\limits_{i=1}^{n} e^{-(\frac{i}{n})^2}}$. Observe that $Q_i^2 = \frac{b^2}{n^2}e^{-\frac{2i^2}{n^2}} < \frac{b^2}{n^2}e^{-\frac{2i^2+2}{n^2}} = Q_{i-1}Q_{i+1}$, we use this Q as it is in some sense roughly the most log concave that it can be and $b \in (1, e)$ is a normalization factor that ensures $\sum\limits_{i=1}^{n} Q_i = 1$. Additionally, $Q_i > Q_j$ for $i < j$. Let the sequence $(j_i, k_i)_{1\leq i\leq\frac{n}{6}}$ be defined by $j_i = 6i - 4$ and $k_i = 6i - 1$.

Let m be the smallest odd integer larger than $C\log n$. Since $n < \frac{1}{48C^{\frac{3}{2}}\epsilon^{\frac{1}{2}}(\log\frac{1}{\epsilon})^{\frac{3}{2}}}$, we have $m < C\log\frac{1}{\epsilon}$. For $1 \leq i \leq \frac{n}{6}$, we take

$$C_i = n^3(Q_{6i-1} - \sqrt{Q_{6i}Q_{6i-2}}) = bn^2e^{\frac{36i^2-12i+1}{n^2}}(1 - e^{-\frac{1}{n^2}})$$

Given $\epsilon > 0$ sufficiently small, we can assume that n and m are bigger than sufficiently large constants. Therefore, we have

$$\frac{1}{2n^2} < 1 - e^{-\frac{1}{n^2}} < \frac{1}{n^2}.$$

Thus, $C_i = \Theta(1)$.

Next, we define the sequence of swapped bins by $(j_i, k_i) = (6i - 4, 6i - 1)$ for $1 \leq i \leq n/6$. Finally, we define A_i and g_i so that F_{yes} and F_{no} are given by $C_i/n^3 - Cm^3\epsilon/n(\cos(\frac{2\pi a}{m}) + \cos(\frac{\pi a}{m}))$ and $C_i/n^3 - Cm^3\epsilon/n(\cos(\frac{2\pi(a+\frac{1}{2})}{m}) + \cos(\frac{\pi a}{m}))$, respectively. It will be useful to compare this to another construction producing the same result. Namely:

$$Q'_a := \begin{cases} Q_a & \text{if } i \not\equiv 1 \pmod 3 \\ Q_a + C_i/n^3 & \text{if } x = 6i - 4 \\ Q_a - C_i/n^3 & \text{if } x = 6i - 1. \end{cases}$$

Then we can likewise construct $D_{yes/no}$ from Q' using the same sequence of (j_i, k_i) and letting $A_i = -Cm^3\epsilon/n$ and $g_i = -\cos(\frac{\pi a}{m})$ for all i. We will switch back and forth between these two interpretations as necessary. We note that the Q'_i are all $\Theta(1/n)$ and therefore the x_{max} in the primed interpretation of the construction is $O(Cm^3\epsilon)$.

We now have some important properties to prove about this construction. Namely that distributions from D_{yes} are log-concave, distributions from D_{no} are far from that and that the two are indistinguishable with few samples. To begin:

Lemma 26. *Any distribution p in D_{yes} is log concave with probability 1.*

Proof. Consider a distribution $p \sim D_{yes}$, given a sample of $\delta_i \in F^i_{yes}$, it moves $(6i-4)$th bin up by δ_i and $(6i-1)$th bin down by δ_i. We note that applying Lemma 17 to the Q' formulation we have that $\delta_i \leq C_i/n^3$. On the other hand, as $n^2 \ll C^{-1}(1/\epsilon)(1/m^3)$, we have that $\delta_i > 0$ for all i. Using this, we can check log-concavity of p at each i based on $i \pmod 6$. In particular,

$$\begin{aligned} p^2_{6i-5} - p_{6i-4}p_{6i-6} &= Q^2_{6i-5} - Q_{6i-6}(Q_{6i-4} + \delta_i) \\ &= Q^2_{6i-5} - Q_{6i-6}Q_{6i-4} - \delta_i Q_{6i-6} \\ &= \Omega(1/n^3) - \Omega(1/n^4) > 0. \end{aligned}$$

We are similarly still log-concave at $6i-3$. We also have that

$$\begin{aligned} p_{6i-1} - \sqrt{p_{6i}p_{6i-2}} &= Q_{6i-1} - \delta_i - \sqrt{Q_{6i}Q_{6i-2}} \\ &= C_i/n^3 - \delta_i \geq 0. \end{aligned}$$

The other locations follow immediately from $\delta_i > 0$ as

$$p^2_{6i-4} - p_{6i-3}p_{6i-5} > Q^2_{6i-4} - Q_{6i-3}Q_{6i-5} > 0,$$

and similarly for $6i-2$ and $6i$. □

In order to show that D_{no} is likely far from log-concave we need a Lemma to allow us to show how far a distribution is from log-concave.

Lemma 27. *For a distribution p over $[n]$ with $p_{3i-2} > p_{3i} > \frac{4}{5}p_{3i-2}$, $p_{3i-1} > \frac{3}{4}p_{3i}$ for all $1 \leq i \leq \frac{n}{3}$, and q any log concave distribution over $[n]$, then $d_{TV}(p,q) \geq \frac{1}{2}\sum_{i=1}^{\frac{n}{3}} \max(0, \sqrt{p_{3i-2}p_{3i}} - p_{3i-1})$.*

Proof. It's sufficient to show that $|p_{3i-2} - q_{3i-2}| + |p_{3i-1} - q_{3i-1}| + |p_{3i} - q_{3i}| \geq \max(0, \sqrt{p_{3i-2}p_{3i}} - p_{3i-1})$ for all $1 \leq i \leq \frac{n}{3}$. For a log concave q, $\sqrt{q_{3i-2}q_{3i}} - q_{3i-1} \leq 0$ must hold for all i. Fix a given i, given p a distribution over $[n]$, we will find such $q_{3i-2}, q_{3i-1}, q_{3i}$ that minimizes $|p_{3i-2} - q_{3i-2}| + |p_{3i-1} - q_{3i-1}| + |p_{3i} - q_{3i}|$ subject to the constraint $q_{3i-1} \geq \sqrt{q_{3i}q_{3i-2}}$. If $q_{3i-2} > p_{3i-2}$, we get a better set of $q_{3i-2}, q_{3i-1}, q_{3i}$ by taking $q_{3i-1} = p_{3i-2}$ and keeping q_{3i-1}, q_{3i} unchanged. Similar reasoning applies to the case of $q_{3i-1} < p_{3i-1}$ or $q_{3i} > p_{3i}$. Therefore, we have that the optimal $q_{3i-2}, q_{3i-1}, q_{3i}$ satisfy $q_{3i-2} \leq p_{3i-2}, q_{3i-1} \geq p_{3i-1}, q_{3i} \leq p_{3i}$, so

$$|p_{3i-2} - q_{3i-2}| + |p_{3i-1} - q_{3i-1}| + |p_{3i} - q_{3i}| = p_{3i-2} - q_{3i-2} + q_{3i-1} - p_{3i-1} + p_{3i} - q_{3i}.$$

Let $\vec{p} = (p_{3i-2}, p_{3i-2}, p_{3i})$ and $\vec{q} = (q_{3i-2}, q_{3i-2}, q_{3i})$, $f : \mathbb{R}^3 \to \mathbb{R}$ is a function where $f(\vec{x}) = \sqrt{x_{3i-2}x_{3i}} - x_{3i-1}$, applying the mean value theorem, we have

$$f(\vec{p}) - f(\vec{q}) = (\vec{p} - \vec{q}) \cdot \nabla f(\vec{x})$$

for some $\vec{x} = (x_1, x_2, x_3)$ between $\vec{p}$ and $\vec{q}$. In particular, it's clear that $q_{3i-2} \leq x_1 \leq p_{3i-2}$, $p_{3i-1} \leq x_2 \leq q_{3i-1}$ and $q_{3i} \leq x_3 \leq p_{3i}$. Expanding the dot product, we can get

$$f(\vec{p}) - f(\vec{q}) = \frac{1}{2}\sqrt{\frac{x_3}{x_1}}(q_{3i-2} - p_{3i-2}) + (p_{3i-1} - q_{3i-1}) + \frac{1}{2}\sqrt{\frac{x_1}{x_3}}(q_{3i} - p_{3i}).$$

Notice that if $q_{3i-2} > \frac{p_{3i-2}}{2}$ and $q_{3i} > \frac{p_{3i}}{2}$, using the relation $p_{3i-2} > p_{3i} > \frac{4}{5}p_{3i-2}$, we have $x_1 \leq p_{3i-2} \leq 2p_{3i} < 4q_{3i} \leq 4x_3$ and $x_3 \leq p_{3i} \leq 2p_{3i-2} < 4q_{3i-2} \leq 4x_1$, we have $f(\vec{p}) - f(\vec{q}) \leq |p_{3i-2} - q_{3i-2}| + |p_{3i-1} - q_{3i-1}| + |p_{3i} - q_{3i}|$ since $\frac{1}{2}\sqrt{\frac{x_3}{x_1}}, \frac{1}{2}\sqrt{\frac{x_1}{x_3}} < 1$.

On the other hand, if $q_{3i-2} \leq \frac{p_{3i-2}}{2}$, by manipulating the relations between $p_{3i-2}, p_{3i-1}, p_{3i}$ in the hypothesis, we have

$$|p_{3i-2}-q_{3i-2}|+|p_{3i-1}-q_{3i-1}|+|p_{3i}-q_{3i}| \geq |p_{3i-2}-q_{3i-2}| \geq \frac{p_{3i-2}}{2} \geq p_{3i-2}-\frac{5}{8}p_{3i} > p_{3i-2}-\frac{5}{6}p_{3i-1} > \sqrt{p_{3i-2}p_{3i}}-p_{3i-1}.$$

Similarly, with the case that $q_{3i} \leq \frac{p_{3i}}{2}$, we can show that

$$|p_{3i-2}-q_{3i-2}|+|p_{3i-1}-q_{3i-1}|+|p_{3i}-q_{3i}| \geq |p_{3i}-q_{3i}| \geq \frac{p_{3i}}{2} = (\frac{5}{4}-\frac{3}{4})p_{3i} \geq \frac{5}{4}p_{3i}-p_{3i-1} > \sqrt{p_{3i-2}p_{3i}}-p_{3i-1}.$$

So in any case we have that

$$|p_{3i-2} - q_{3i-2}| + |p_{3i-1} - q_{3i-1}| + |p_{3i} - q_{3i}| \geq \sqrt{p_{3i-2}p_{3i}} - p_{3i-1}.$$

Summing over i yields our result. □

Using this we show that a distribution from D_{no} is likely far from log-concave.

Lemma 28. *With 99% probability, a random distribution drawn from D_{no} is ϵ far from log concave.*

Proof. Let p be taken from D_{no} and let q be an arbitrary log-concave distribution over $[n]$. Applying Lemma 17, we note that for each i there is independently a $1/m$ probability that $\delta_i = C_i/n^3 + \Omega(Cm\epsilon/n)$. Let X be the number of such i's. We note that for each such i that

$$\begin{aligned} p_{6i-1} - \sqrt{p_{6i}p_{6i-2}} = & \\ & = Q_{6i-1} - \sqrt{Q_{6i}Q_{6i-2}} - \delta_i \\ & = C_i/n^3 - \delta_i < -\Omega(Cm\epsilon/n). \end{aligned}$$

Therefore, by Lemma 27, $d_{TV}(p, q) \geq \Omega(XCm\epsilon/n)$. It thus suffices to show that with 99% probability that $X = \Omega(n/m)$. However, as $X \sim \text{Bin}(n/6, 1/m)$, this is clear. □

Finally, we can complete our proof of Theorem 5.

Proof. We begin by proving it for n less than $\frac{1}{C^2\epsilon^{\frac{1}{2}}(\log\frac{1}{\epsilon})^{\frac{3}{2}}}$ and a multiple of 6. Suppose that there is a tester that can reliably distinguish between a log-concave distribution and one that is ϵ-far using $N < C^{-3}n\epsilon^{-2}\log^{-7}(1/\epsilon)$ samples. As a distribution from D_{yes} is log-concave and one from D_{no} is likely ϵ-far our tester can reliably distinguish the two. However, using the Q' interpretation of our construction, we have $B = O(N/n)$ so $Bx_{max}^2 = O(C^2m^6\epsilon^2N/n) = O(C^{-1}/\log(n/\epsilon x_{max}))$. Thus, we can apply Corollary 10 to see that $d_{TV}(D_{yes}^N, D_{no}^N) < 1/100$ which contradicts our algorithm being able to reliably distinguish them.

For other n, we can just apply this result to n_0, the largest integer that satisfies our conditions. As a log-concave distribution on $[n_0]$ is also a log-concave distribution over $[n]$, this gives a reduction between the testing problems and completes the proof. □

6 Conclusion

In this paper we have produced a general framework for proving distribution testing lower bounds for properties defined by local inequalities between the individual bin probabilities. Applying it requires finding an instantiation of our construction so that many bins satisfy these inequalities tightly and changing their values slightly will break the property in question. Usually, this technique should give lower bounds comparable to the testing-by-learning algorithm of $O(n/\epsilon^2)$ samples up to logarithmic factors so long as n is not too big, while for larger values of n it will often fail to find further improvements.

As applications of this new technique we have proved new lower bounds for monotonicity testing, and nearly optimal lower bounds for log-concavity testing.

References

[1] J. Acharya, C. Daskalakis, and G. Kamath *Optimal testing for properties of distributions* in Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7–12, 2015, Montreal, Quebec, Canada, pages 3591—3599, 2015.

[2] T. Batu, R. Kumar, and R. Rubinfeld *Sublinear algorithms for testing monotone and unimodal distributions* in Proceedings of the Thirty-sixth Annual ACM Symposium on Theory of Computing, STOC '04, pages 381-–390, New York, NY, USA, 2004. ACM.

[3] A. Bhattacharyya, E. Fischer, R. Rubinfeld, and P. Valiant *Testing monotonicity of distributions over general partial orders* in ICS, pages 239—252, 2011.

[4] C. Canonne *A Survey on Distribution Testing: Your Data is Big. But is it Blue?* Theory of Computing, Graduate Surveys 9:1–100, 2020.

[5] C. Canonne, I. Diakonikolas, T. Gouleakis, and R. Rubinfeld *Testing shape restrictions of discrete distributions* Theory Comput. Syst., 62(1):4—62, 2018.

[6] C. Canonne, I. Diakonikolas, A. Stewart *Testing for Families of Distributions via the Fourier Transform*

[7] J. Neyman and E. S. Pearson *On the problem of the most efficient tests of statistical hypotheses* Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character, 231(694-706):289-–337, 1933.

[8] R. Rubinfeld and M. Sudan *Robust characterizations of polynomials with applications to program testing* SIAM J. on Comput., 25:252—271, 1996.