Testing Unate Distributions

Daeho Lee
Massachusetts Institute of Technology, Cambridge, MA, USA

Shivam Nadimpalli
Massachusetts Institute of Technology, Cambridge, MA, USA

Mingda Qiao
University of Massachusetts Amherst, MA, USA

Ronitt Rubinfeld
Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

We initiate the study of *unate distributions* over $\{\pm 1\}^n$ – a natural analogue of unate Boolean functions – by considering two basic testing problems that parallel well-studied questions for monotone distributions:

- **Uniformity Testing of Unate Distributions:** We show that $\widetilde{\Theta}(n^{3/2})$ samples are sufficient and necessary, in contrast to the $\widetilde{\Theta}(n)$ sample complexity of the analogous problem for monotone distributions (Rubinfeld and Servedio, STOC 2005; Adamaszek, Czumaj, and Sohler, SODA 2010).
- **Unateness Testing of Arbitrary Distributions:** We give a tester that uses $\widetilde{O}(n^{3/2})$ conditional samples in the subcube conditional model. On the other hand, every tester that draws conditional samples in a similar fashion, namely from $O(1)$-dimensional subcubes, must have an $\widetilde{\Omega}(n^{2/3})$ complexity. In the same model, the complexity of monotonicity testing was recently shown to be $\widetilde{\Theta}(n)$ (Chakrabarty et al., STOC 2025).

Our algorithms for both problems significantly outperform the naive approach of reducing to the monotone case, which would incur $\Omega(n^2)$ sample complexity. Our uniformity tester relies on a subroutine that "weakly" learns the hidden orientations of a unate distribution, together with a new correlation bound for these estimates. Both tools may be of independent interest in studying monotonicity and unateness over $\{\pm 1\}^n$.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms; Mathematics of computing → Probability and statistics

Keywords and phrases Distribution testing, unate distributions, monotone distributions, uniformity testing, subcube conditioning

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2026.57

Category RANDOM

Related Version *Full Version*: `https://arxiv.org/abs/2607.01573` [45]

Funding *Daeho Lee*: Supported by the MIT Undergraduate Research Opportunities Program (UROP).
Shivam Nadimpalli: Partially supported by Elchanan Mossel's Vannevar Bush Faculty Fellowship ONR-N00014-20-1-2826.
Ronitt Rubinfeld: Supported by the NSF TRIPODS program (award DMS-2022448) and CCF-2310818.

1 Introduction

Many fundamental distribution testing tasks – uniformity testing being the canonical example – become infeasible in high dimensional settings such as the Boolean hypercube $\{\pm 1\}^n$, requiring exponentially many samples in the dimension [15, 18]. A central theme in the field has therefore been to identify *structured* settings that admit efficient testers: examples include Bayesian networks [17, 35], Markov random fields [34, 8], and structured truncations [36, 37].

Among such structural assumptions, one of the most extensively studied is *monotonicity*, the distributional analogue of the classical Boolean function property.[1] In the Boolean function setting, monotonicity has played a central role in sublinear algorithms for nearly three decades: it was among the first properties studied in Boolean function testing and continues to be actively investigated to this day [14, 38, 22, 29, 41, 43, 26, 11, 44, 25]. In the distributional setting, monotonicity has similarly been studied in depth, leading to tight upper and lower bounds for tasks such as uniformity testing and property testing under various access models [6, 51, 2, 9, 1, 4, 52, 12, 21].

A closely related property in the Boolean function setting is *unateness*. Informally, a Boolean function is unate if each coordinate is either always non-decreasing or always non-increasing. The problem of testing unateness of Boolean functions was introduced alongside monotonicity testing in [38], and has since received considerable attention [42, 23, 31, 32, 30, 5]. Both monotonicity and unateness are fundamental structural properties that arise naturally across many domains, for example in social choice (where we can view a Boolean function f as a voting rule), economics, learning theory, and circuit design.

Despite extensive work on unate functions, the analogous notion for distributions has, to the best of our knowledge, not been studied. In this work, we initiate the study of *unate distributions* over $\{\pm 1\}^n$. The motivation for this is twofold:

- The study of monotone distributions has a well-developed literature precisely because monotonicity is a natural structural constraint arising in settings ranging from social choice to learning theory. Unateness is a strict and natural generalization – it captures the same "directional" structure without fixing a canonical orientation – and we believe its study for distributions will prove equally fruitful.
- Additionally, unate distributions, which we define shortly, form a broad and natural family of probability distributions. For example, every unate function f induces a unate distribution (the uniform distribution over $f^{-1}(1)$), and the class of unate functions includes families such as halfspaces and read-once decision lists, which need not be monotone. Beyond this connection to functions, every product distribution over $\{\pm 1\}^n$ is unate.

Finally, as this work demonstrates, unate distributions pose challenges that cannot be addressed by simply reducing to the monotone setting. Even for basic tasks such as uniformity testing, the sample complexity changes, and more broadly, new ideas are required to design and analyze optimal testers for this richer class of distributions.

1.1 Our Results

We start by formally defining unate distributions over $\{\pm 1\}^n$:

▶ **Definition 1.** *A distribution $\mathcal{D}$ over $\{\pm 1\}^n$ is* unate *if there exists $\sigma \in \{\pm 1\}^n$ such that $\mathcal{D}(x \oplus \sigma)$ is a monotone probability mass function, where $\oplus$ denotes coordinate-wise multiplication.*

[1] Formally, a distribution $\mathcal{D}$ over $\{\pm 1\}^n$ is *monotone* if its probability mass function $\mathcal{D} : \{\pm 1\}^n \to [0, 1]$ is monotone, i.e., $\mathcal{D}(x) \leq \mathcal{D}(y)$ whenever $x_i \leq y_i$ for all $i \in [n]$.

Unate distributions immediately inherit many of the algorithmic challenges and lower bounds known from the monotone distribution setting. This includes, for example, the exponential sample lower bounds for tasks such as entropy estimation and independence testing [51]. In this work, we give efficient algorithms as well as lower bounds for two basic problems:

- Uniformity testing of unate distributions; and
- Testing unateness of an unknown distribution in the *subcube conditional model* introduced by Canonne, Ron, and Servedio [20] and Chakraborty, Fischer, Goldhirsh, and Matsliah [24].

In the subcube conditional model for distributions over $\{\pm 1\}^n$, an algorithm can query $\mathcal{D}$ with a subcube $\rho \in \{\pm 1, *\}^n$ and receive an independent sample $\boldsymbol{x} \sim \mathcal{D}$ conditioned on $\boldsymbol{x}_i = \rho_i$ for all i where $\rho_i \neq *$. This model provides a distributional analogue of the "membership query" from learning theory and property testing of Boolean functions, and has received considerable attention in recent years.

We note that for both of the problems we consider, an easy argument using the Chernoff bound together with a union bound over all $\sigma \in \{\pm 1\}^n$ (cf. Definition 1) allow us to reduce to the corresponding problems for monotone distributions with a multiplicative $O(n)$ sample/subcube query overhead. Our main algorithmic results improve on this naive approach for each task; we defer a technical overview of our results to Section 1.2 and survey related work in Section 1.3.

1.1.1 Uniformity Testing of Unate Distributions

Throughout, we write $\mathcal{U} = \mathcal{U}_n$ for the uniform distribution over $\{\pm 1\}^n$. Our first result gives an efficient uniformity tester for high-dimensional distributions under the promise of unateness:

▶ **Theorem 2.** *Let $\varepsilon \in (0,1)$. There is an algorithm,* `Unate-Uniformity` *(Algorithm 1) which, given i.i.d. sample access to an unknown unate distribution $\mathcal{D}$ over $\{\pm 1\}^n$, draws $\widetilde{O}(n^{3/2}/\varepsilon^2)$ samples and has the following performance guarantee:*

- *If $\mathcal{D} = \mathcal{U}$, then it outputs "accept" with probability $9/10$.*
- *If $\mathrm{d_{TV}}(\mathcal{D},\mathcal{U}) \geq \varepsilon$, then it outputs "reject" with probability $9/10$.*

Furthermore, it runs in $\widetilde{O}(n^{5/2}/\varepsilon^2)$ time.

Here $\mathrm{d_{TV}}(\cdot)$ refers to the total variation or statistical distance (Section 2.2). This result should be contrasted with the $\widetilde{O}(n/\varepsilon^2)$ bound for uniformity testing of monotone distributions due to Rubinfeld and Servedio [51].[2] At a high level, moving from monotone to unate distributions introduces an unknown orientation vector σ (cf. Definition 1); any naive strategy that first learns σ and then runs the [51] tester provably incurs an $\Omega(n^2)$ sample overhead. Our algorithm bypasses this bottleneck by never learning σ explicitly. Instead, we introduce a weak-orientation learning framework whose errors are provably weakly correlated across coordinates. This allows a small average bias to be amplified into a global signal using only $\widetilde{O}(n^{3/2})$ samples. We believe this structural result on limited correlations among marginals of unate distributions is of independent interest. We complement Theorem 2 with a matching lower bound:

[2] See also Adamaszek, Czumaj, and Sohler [2] who improved the sample complexity to $O(n/\varepsilon^2)$.

▶ **Theorem 3.** *Let $\mathcal{A}$ be any algorithm which, given i.i.d. sample access to an unknown distribution $\mathcal{D}$, has the performance guarantee from Theorem 2 with $\varepsilon = 1 - n^{-10}$. Then $\mathcal{A}$ must draw $\Omega(n^{3/2}/\log^2 n)$ samples from $\mathcal{D}$.*

Our proof of Theorem 3 extends the "monotone decomposition method" of Rubinfeld and Servedio [51] to the setting of unate distributions; see Section 1.2 for more details.

▶ Remark 4. One might hope that the additional power of the subcube conditional model could yield improved bounds for uniformity testing of unate distributions. Recall that uniformity testing of *arbitrary* distributions can be done using $\widetilde{O}(\sqrt{n})$ subcube queries [16]. Unfortunately, it turns out that the additional structure of unateness cannot beat this baseline: Chakrabarty, Chen, Ristic, Seshadhri, and Waingarten [21] have recently shown that even the easier problem of testing uniformity of monotone distributions in the subcube model already requires $\widetilde{\Omega}(\sqrt{n})$ queries.

1.1.2 Testing Unateness of Distributions

We now turn to the problem of testing whether an unknown distribution over $\{\pm 1\}^n$ is itself unate. In the standard access model where one receives i.i.d. samples from the unknown distribution being tester, this task is intractable: even for the simpler case of monotonicity, classical "birthday paradox" arguments imply that exponential-in-n sample complexity is unavoidable [4, 52]. It is therefore necessary to consider stronger access models to obtain meaningful algorithms.

Motivated by this, Chakrabarty, Chen, Ristic, Seshadhri, and Waingarten [21] studied monotonicity testing in the subcube conditional model, where the tester can query samples conditioned on arbitrary coordinate restrictions. Building on their results, we give an efficient algorithm for testing unateness in the same model.

▶ **Theorem 5.** *Let $\varepsilon \in (0, 1)$. There is an algorithm,* `Subcube-Unate` *(Algorithm 2) which, given subcube query access to an unknown distribution $\mathcal{D}$ over $\{\pm 1\}^n$, makes $\widetilde{O}(n^{3/2}/\varepsilon^2)$ queries and has the following performance guarantee:*

- *If $\mathcal{D}$ is unate, it outputs "accept" with probability at least $2/3$.*
- *If $\mathcal{D}$ is ε-far in TV distance from every unate distribution on $\{\pm 1\}^n$, then it outputs "reject" with probability at least $2/3$.*

We note that the algorithm of Theorem 5 also works in the weaker *coordinate oracle* model of Blanca et al. [13] where queries are only made on one-dimensional subcubes. We complement Theorem 5 with the following lower bound:

▶ **Theorem 6.** *There exists a constant $\varepsilon_0 > 0$ such that the following holds: If an algorithm tests whether an unknown distribution $\mathcal{D}$ over $\{\pm 1\}^n$ is unate or ε_0-far from unate by drawing Q_1 samples from $\mathcal{D}$ and making Q_2 (possibly adaptive) subcube queries on subcubes of dimensions $\leq d$, we have*

$$Q_1 + 2^d Q_2 \geq \Omega(n^{2/3}/\log^3 n).$$

We note that all prior lower bounds in the subcube model are based on product distributions as hard instances. Since every product distribution is trivially unate, these techniques cannot be applied to this setting. To the best of our knowledge, ours is the first lower bound that breaks this barrier; we return to this point in Section 1.4.

1.2 Main Ideas

We give a brief technical overview of our results. Although both of our upper bounds end up with the same asymptotic sample complexity of $\widetilde{O}(n^{3/2}/\varepsilon^2)$, that the ideas underlying each of them are quite different.

1.2.1 Testing Uniformity of Unate Distributions

Upper Bound. A natural first attempt is to adapt the Rubinfeld–Servedio [51] uniformity tester for monotone distributions to the unate setting. The idea would be to first learn the hidden "orientation vector" $\sigma^* \in \{\pm 1\}^n$ (cf. Definition 1) such that the reoriented distribution $\mathcal{D}^{\sigma^*}$, where $\boldsymbol{x} \sim \mathcal{D}^{\sigma^*}$ is obtained by drawing $\boldsymbol{y} \sim \mathcal{D}$ and setting $\boldsymbol{x}_i = \sigma_i^* \cdot \boldsymbol{y}_i$, is monotone. Since total variation distance is preserved under bijections, $\mathrm{d_{TV}}(\mathcal{D},\mathcal{U}) = \mathrm{d_{TV}}(\mathcal{D}^{\sigma^*},\mathcal{U})$. Thus, if σ^* were known, one could directly apply the $O(n/\varepsilon^2)$-sample uniformity tester of [51, 2].

However, exactly recovering σ^* can be prohibitively expensive. Consider, for example, a product distribution where each coordinate has mean $\pm\Theta(1/n)$. Such a distribution is $\Omega(1)$-far from uniform, but distinguishing $\mathbf{E}[x_i] = 0$ from $\mathbf{E}[x_i] = \pm\Theta(1/n)$ for each coordinate requires $\Omega(n^2)$ samples by Chernoff bounds. This naive approach of fully learning the orientation thus fails to give the desired sample complexity of $\widetilde{O}(n^{3/2})$.

Our algorithm circumvents this barrier by avoiding the need to learn σ^* exactly. Instead, we estimate a weakly correlated proxy for σ^* using only $\widetilde{O}(n^{3/2}/\varepsilon^2)$ samples. The key technical step (Lemma 11; which may be of independent interest) shows that for unate distributions, the coordinate-wise sign estimates cannot exhibit strong correlations: indeed, their covariances decay at rate $O(1/\sqrt{m})$ after m samples. This weak dependence is sufficient to amplify a small average bias across coordinates into a global signal, allowing us to test uniformity with $\widetilde{O}(n^{3/2})$ samples.

Lower Bound. The matching lower bound for uniformity testing is proved in the full version [45].

1.2.2 Testing Unate Distributions via Subcube Queries

Upper Bound. For the task of testing whether an unknown distribution is itself unate, we work in the *subcube conditional model*, following the recent work of Chakrabarty et al. [21]. At a high level, our algorithm adapts the classical *edge tester* for Boolean unateness [38] to the distributional setting. The central tool is a structural lemma of [21], based on a real-valued "directed" generalization of an isoperimetric inequality due to Talagrand [53], which controls the "edge bias" of monotone distributions. We refer the reader to Lemma 14 for a precise statement. Using this lemma, we show that if $\mathcal{D}$ is ε-far from unate, then there exists some coordinate i that simultaneously witnesses both "monotone" and "anti-monotone" violations. This can then be detected by making $\widetilde{O}(n^{3/2})$ subcube edge queries; the analysis requires an elementary but careful calculation.

Lower Bound. The lower bound for unateness testing in the subcube model is proved in the full version [45].

1.3 Related Work

Monotonicity of Boolean functions is among the most extensively studied properties in sublinear algorithms; we will not attempt to survey the very large body of results here and instead refer the reader to the comprehensive discussion from [21]. The problem of unateness

testing of Boolean functions was introduced alongside that of monotonicity testing in [38], and a sequence of works [38, 42, 23, 31, 32, 30] have pinned down its exact query complexity: $\widetilde{\Theta}(n^{2/3})$ queries are both necessary and sufficient to test unateness of Boolean functions.

A number of natural distribution testing tasks have exponential sample complexity in high-dimensional settings. This includes uniformity testing over $\{\pm 1\}^n$; see the surveys [15, 18] for more discussion on this. In order to overcome these lower bounds, many works either make structural assumptions on the distribution being tested, or assume stronger access models to the distribution. Examples of the former include monotonicity [51, 2], Bayesian networks [17, 35], Markov random fields [34, 8], various classes of structured truncations [40, 36, 37]; see also Section 7 of [15].

The subcube conditioning model, introduced by [20, 24, 10], takes the other route of assuming stronger access to the distribution being tested. This model has received much attention in recent years [16, 27, 47, 28, 3, 13, 21], and has also found applications beyond distribution testing to problems in learning theory [27, 12]. There has also been a nascent line of work on using the subcube conditioning model to make high-dimensional distribution testing practical [46, 49].

The works most relevant to our results are [51] and [21]. Rubinfeld and Servedio [51] establish an essentially tight bound of $\widetilde{\Theta}(n)$ samples for the problem of testing uniformity of an unknown monotone distribution, whereas Charkabarty et al. [21] give $\widetilde{\Theta}(n)$-query upper and lower bounds for testing monotonicity of an unknown distribution in the subcube model. Both our results – for uniformity testing as well as unateness testing – rely on ingredients going into the analyses of [51, 21] respectively.

1.4 Discussion

Our results raise a number of intriguing questions for further study; we highlight two compelling directions below.

Lower Bounds for Unateness Testing with Subcube Queries. Recall that Theorem 6 gives a $\widetilde{\Omega}(n^{2/3})$ query lower bound against algorithms that make $O(1)$-dimensional subcube queries. At present, all known lower-bounds in the subcube conditional model [16, 27, 21] construct hard instances that are product distributions, which are amenable to moment-matching techniques [27, 21]. However, since product distributions are automatically unate, none of these techniques can be lifted the unateness setting. Improving on Theorem 6 is thus likely to lead to new insights and techniques for proving lower bounds in the subcube conditional model.

***k*-Monotone Distributions.** Finally, one can consider distributional analogues of other structural generalizations of monotonicity. A natural candidate is the class of *k-monotone distributions*, where the distribution's mass function changes direction at most k times along any monotone path from -1^n to 1^n. The class of k-monotone functions has already been studied in the Boolean function setting [39, 19, 33], and k-monotone distributions could provide another structured setting for distribution testing beyond monotonicity.

2 Preliminaries

We use boldfaced letters such as (e.g. $\boldsymbol{x} \sim \{\pm 1\}^n$) to denote random variables. Unless explicitly stated otherwise, all probabilities and expectations will be with respect to the uniform distribution. Throughout, we will write $\mathcal{U}_n$ for the uniform distribution over $\{\pm 1\}^n$ and when the dimension n is clear from context we will simply write $\mathcal{U}$ instead.

For a distribution $\mathcal{D}$, it will be convenient to write $\mathcal{D}^{\otimes m}$ for the distribution of m independent draws from $\mathcal{D}$. We will use the following notation throughout:

▶ **Notation 7.** *Given a distribution $\mathcal{D}$ over $\{\pm 1\}^n$ and $\sigma \in \{\pm 1\}^n$, we write $\mathcal{D}^\sigma$ for the distribution over $\{\pm 1\}^n$ where a draw $\boldsymbol{x} \sim \mathcal{D}^\sigma$ is obtained by first drawing $\boldsymbol{y} \sim \mathcal{D}$ and then setting $\boldsymbol{x}_i \coloneqq \sigma_i \cdot \boldsymbol{y}_i$ for $i \in [n]$.*

2.1 The Berry–Esseen Theorem

We will write $N(0, I_d)$ for the d-dimensional standard Gaussian distribution, where I_d denotes the $d \times d$ identity matrix. We recall the (multi-dimensional) Berry–Esseen theorem (which will be used to prove Lemma 11):

▶ **Theorem 8** (Theorem 1.1 of [7]). *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_m$ be i.i.d. random vectors in $\mathbb{R}^d$ distributed as $\boldsymbol{X}$ where $\mathbf{E}[\boldsymbol{X}] = 0$ and $\mathbf{E}[\boldsymbol{X}\boldsymbol{X}^\top] = I_d$. Then for any convex set $A \subseteq \mathbb{R}^d$, we have*

$$\left| \mathbf{Pr}\left[\frac{\sum_{i=1}^m \boldsymbol{X}_i}{\sqrt{m}} \in A \right] - \mathbf{Pr}[\boldsymbol{G} \in A] \right| \leq O(1) \cdot \frac{d^{1/4}}{\sqrt{m}} \cdot \mathbf{E}\left[\|\boldsymbol{X}\|^3 \right].$$

where $\boldsymbol{G} \sim N(0, I_d)$.

2.2 Distance Metrics Between Probability Distributions

We will identify a distribution $\mathcal{D}$ over a discrete domain Ω with its density function $\mathcal{D} : \Omega \to [0, 1]$. Recall that for two distributions $\mathcal{D}_1$, $\mathcal{D}_2$ over a discrete domain Ω, the *total variation* (or *statistical*) distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is

$$\mathrm{d}_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \sum_{x \in \Omega} \left| \mathcal{D}_1(x) - \mathcal{D}_2(x) \right|.$$

We will frequently make use of the fact that total variation distance is invariant under bijections.

We also recall the *Kullback–Leibler (KL)* and χ^2-divergences:

$$\mathrm{d}_{\mathrm{KL}}(\mathcal{D}_1, \mathcal{D}_2) = \sum_{x \in \Omega} \mathcal{D}_1(x) \log \frac{\mathcal{D}_1(x)}{\mathcal{D}_2(x)} \qquad \text{and} \qquad \mathrm{d}_{\chi^2}(\mathcal{D}_1, \mathcal{D}_2) = \sum_{x \in \Omega} \frac{\left(\mathcal{D}_1(x) - \mathcal{D}_2(x)\right)^2}{\mathcal{D}_2(x)}$$

respectively. The following relationships are standard:

$$\mathrm{d}_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) \leq \sqrt{\frac{1}{2} \mathrm{d}_{\mathrm{KL}}(\mathcal{D}_1, \mathcal{D}_2)} \leq \sqrt{\frac{1}{2} \log\left(1 + \mathrm{d}_{\chi^2}(\mathcal{D}_1, \mathcal{D}_2)\right)}. \tag{1}$$

We will use the following version of the data processing inequality:

▶ **Fact 9.** *Let $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ be random variables over the same domain. For any (possibly randomized) algorithm $\mathcal{A}$, we have $\mathrm{d}_{\mathrm{TV}}\big(\mathcal{A}(\boldsymbol{X}_1), \mathcal{A}(\boldsymbol{X}_2)\big) \leq \mathrm{d}_{\mathrm{TV}}(\boldsymbol{X}_1, \boldsymbol{X}_2)$.*

3 Testing Uniformity of Unate Distributions

We will prove Theorems 2 and 3 in this section, starting with the former.

3.1 Upper Bound

We first recall the principal technical lemma of [51]'s uniformity tester for monotone distributions:

▶ **Lemma 10** (Corollary 6 of [51]). *If $\mathcal{D}$ is a monotone distribution over $\{\pm 1\}^n$ that is ε-far from uniform, then*

$$\mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}}\left[\sum_{i=1}^{n}\boldsymbol{x}_i\right] \geq \varepsilon\,.$$

We now turn to our uniformity tester for unate distributions. First, note that if $\mathcal{D}$ is a unate distribution, then there exists some $\sigma^* \in \{\pm 1\}^n$ such that $\mathcal{D}^{\sigma^*}$ is a monotone distribution. In particular, note that we may take $\sigma_i^* = \mathrm{sign}(\mathbf{E}_{\boldsymbol{x}\sim\mathcal{D}}[\boldsymbol{x}_i])$. This suggests a natural algorithm: estimate σ^* by drawing m samples $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(m)}$ from $\mathcal{D}$ and setting

$$\boldsymbol{\sigma}_i := \mathrm{sign}\left(\sum_{j=1}^{m}\boldsymbol{x}_i^{(j)}\right),$$

and then run the [51] uniformity tester on the monotone distribution $\mathcal{D}^{\boldsymbol{\sigma}}$. Indeed, since

$$\mathrm{d}_{\mathrm{TV}}(\mathcal{D}^{\boldsymbol{\sigma}},\mathcal{U}) = \mathrm{d}_{\mathrm{TV}}(\mathcal{D}^{\boldsymbol{\sigma}},\mathcal{U}^{\boldsymbol{\sigma}}) = \mathrm{d}_{\mathrm{TV}}(\mathcal{D},\mathcal{U}),$$

replacing $\mathcal{D}$ with $\mathcal{D}^{\boldsymbol{\sigma}}$ preserves the uniformity (or distance from uniformity) of the distribution.

However, learning the whole sign vector σ^* can be too expensive. In particular, while Lemma 10 ensures that $\sum_{i=1}^{n}\mathbf{E}[\boldsymbol{x}_i]$ is large, it gives no guarantee on the magnitude of each individual $\mathbf{E}[\boldsymbol{x}_i]$. In particular, note that distinguishing $\mathbf{E}[\boldsymbol{x}_i] = 0$ (which corresponds to the uniform distribution) from $\mathbf{E}[\boldsymbol{x}_i] = \pm\Theta(1/n)$ requires $\Theta(n^2)$ samples per coordinate by standard Chernoff lower bounds. We bypass this obstacle by not insisting on the exact orientation and instead *weakly* learning the sign vector σ^*.

3.1.1 Warm-Up: A Sub-Quadratic Tester in a Special Case

As a warm-up, we show how to obtain a sub-quadratic tester in the hard instance just described above, where naively learning σ^* requires $\Omega(n^2)$ queries. It will be convenient to write

$$\mu_i := \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}}[\boldsymbol{x}_i]\,.$$

We make one simplifying assumption for now: $\mu_i = \frac{1}{n}$ holds for every $i \in [n]$, i.e., the unate distribution $\mathcal{D}$ is in fact monotone, and each coordinate has the same bias of $\Theta(1/n)$ towards $+1$.

Let $p > 0$ be a parameter that we will set later. By the assumption that $\mu_i = \frac{1}{n}$, it can be verified that $m = \Theta(n^2p^2)$ samples suffice for the "orientation step" to produce random signs $\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_n \in \{\pm 1\}$ with the promise that

$$\mathbf{E}[\boldsymbol{\sigma}_i] \geq p \text{ for all } i \in [n]\,.$$

Equivalently, we correctly deduce that $\sigma_i^* = +1$ with probability $1/2 + \Omega(p)$, which is slightly better than a random guess.

Writing $\boldsymbol{X} \coloneqq \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i \in [-1,1]$, note that $\mathbf{E}[\boldsymbol{X}] \geq p$. Applying Markov's inequality to the non-negative random variable $1 - \boldsymbol{X}$ gives

$$\mathbf{Pr}\left[\boldsymbol{X} \leq \frac{p}{2}\right] = \mathbf{Pr}\left[1 - \boldsymbol{X} \geq 1 - \frac{p}{2}\right] \leq \frac{\mathbf{E}\left[1-\boldsymbol{X}\right]}{1-\frac{p}{2}} \leq \frac{1-p}{1-\frac{p}{2}} \leq 1 - \Omega(p)\,.$$

In other words, the event $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i \geq \frac{p}{2}$ holds with probability $\Omega(p)$. Repeating this "orientation step" independently $O(1/p)$ times ensures that with constant probability, at least one draw satisfies $\boldsymbol{X} \geq p/2$. Then, we note that, conditioning on the realization of $\boldsymbol{\sigma} \in \{\pm 1\}^n$, the $\boldsymbol{\sigma}$-weighted Hamming weight $\sum_{i=1}^{n} \boldsymbol{\sigma}_i \boldsymbol{x}_i$ for $\boldsymbol{x} \sim \mathcal{D}$ has an expectation of exactly

$$\sum_{i=1}^{n} \boldsymbol{\sigma}_i \cdot \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}}\left[\boldsymbol{x}_i\right] = \sum_{i=1}^{n} \boldsymbol{\sigma}_i \cdot \mu_i = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i.$$

Therefore, by estimating $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i$ to an additive error of $O(p)$ using $\widetilde{O}(n/p^2)$ additional samples, we obtain a tester with sample complexity (modulo polylogarithmic factors)

$$\frac{1}{p} \cdot n^2 p^2 + \frac{n}{p^2}. \tag{2}$$

Setting $p = n^{-1/3}$ gives a sample complexity of $\widetilde{O}(n^{5/3})$, which already improves upon the easy $\widetilde{O}(n^2)$ bound (cf. Section 1.1).

3.1.2 Main Technical Lemma

Recall from the warm-up that the sign estimates $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2, \ldots, \boldsymbol{\sigma}_n$ satisfy $\mathbf{E}\left[\boldsymbol{\sigma}_i\right] \geq p$, which implies

$$\mathop{\mathbf{Pr}}_{\boldsymbol{\sigma}}\left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i \geq \Omega(p)\right] \geq \Omega(p)\,.$$

Now, note that if we were able to replace the R.H.S. above with $\Omega(1)$ instead of $\Omega(p)$, then this stronger lower bound would imply that $O(1)$ independent sign estimates would suffice for "catching" the $\Omega(p)$ bias. Then, the sample complexity (up to polylogarithmic factors) would be $n^2p^2 + \frac{n}{p^2}$ (cf. Equation (2)), which reduces to $\widetilde{O}(n^{3/2})$ if we set $p = n^{-1/4}$.

We first note that the property $\mathbf{E}\left[\boldsymbol{\sigma}_i\right] \geq p$ alone does not imply the desired high-probability bound, namely,

$$\mathbf{Pr}\left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i \geq \Omega(p)\right] \geq \Omega(1).$$

Suppose for the sake of simplicity that n is even, and imagine that $\boldsymbol{\sigma} \in \{\pm 1\}^n$ is drawn from the following distribution:

- With probability p, $\boldsymbol{\sigma} = (+1, +1, \ldots, +1)$.
- With the remaining probability $1-p$, $\boldsymbol{\sigma}$ is chosen as a uniformly random permutation of $n/2$ copies of $+1$ and $n/2$ copies of -1.

It can be easily verified that the above distribution satisfies $\mathbf{E}\left[\boldsymbol{\sigma}_i\right] = p$, yet $\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\sigma}_i > 0$ only holds with probability p.

Therefore, the crux is to show that the $\boldsymbol{\sigma}_i$'s cannot be too correlated (as in the example above). Concretely, proving the following will suffice for our purposes:

Algorithm 1 The `Unate-Uniformity` algorithm.

Input: Access to i.i.d. samples from unate $\mathcal{D}$, distance parameter $\varepsilon \in (0,1)$

Output: "Accept" or "reject"

`Unate-Uniformity`$(\mathcal{D}, \varepsilon)$:

1. Repeat the following $O(1)$ times:
   a. Draw $m_1 := \Theta\left(\frac{n^{3/2}}{\varepsilon^2}\right)$ samples $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(m_1)} \sim \mathcal{D}$.
   b. Compute $\boldsymbol{\sigma} \in \{\pm 1\}^n$ where for $i \in [n]$ we have

$$\boldsymbol{\sigma}_i := \operatorname{sign}\left(\sum_{j=1}^{m_1} \boldsymbol{x}_i^{(j)}\right) \in \{\pm 1\}\,.$$

   c. Draw $m_2 := \Theta\left(\frac{n^{3/2}}{\varepsilon^2} \log\left(\frac{n}{\varepsilon}\right)\right)$ samples $\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(m_2)} \sim \mathcal{D}$.
   d. For each $j \in [m_2]$, let $\boldsymbol{w}_j := \sum_{i=1}^n \boldsymbol{\sigma}_i \cdot \boldsymbol{y}_i^{(j)}$ denote the $\boldsymbol{\sigma}$-weighted Hamming weight of $\boldsymbol{y}^{(j)}$.
   e. If $|\boldsymbol{w}_j| \geq \Theta(\sqrt{n \log(n/\varepsilon)})$ holds for any $j \in [m_2]$, halt and output "reject."
   f. If $\frac{1}{m_2} \sum_{j=1}^{m_2} \boldsymbol{w}_j \geq \Theta(\varepsilon/n^{1/4})$, then halt and output "reject."
2. If the algorithm has not rejected yet, output "accept."

▶ **Lemma 11.** *Suppose $\mathcal{D}$ is a unate distribution on $\{\pm 1\}^n$. Suppose $\boldsymbol{x}^{(1)}, \ldots \boldsymbol{x}^{(m)} \sim \mathcal{D}$ and let*

$$\boldsymbol{\sigma}_i := \operatorname{sign}\left(\sum_{j=1}^{m} \boldsymbol{x}_i^{(j)}\right).$$

For any $1 \leq i < j \leq n$, it holds that

$$|\operatorname{\mathbf{Cov}}[\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j]| = O\left(\frac{1}{\sqrt{m}}\right).$$

It is readily verified that Lemma 11 fails for non-unate distributions: consider, for example, the two-point distribution $\mathcal{D}(+1^n) = \mathcal{D}(-1^n) = 0.5$. We defer the proof of Lemma 11 to Section 3.1.4, and first show why it implies the correctness of Algorithm 1.

3.1.3 Proof of Theorem 2

We will require the following lemma due to Qiao and Valiant [50]:

▶ **Lemma 12** (Lemma B.2 of [50])**.** *Suppose $k \in \mathbb{N}$ and $\delta \in (0, 1/15\sqrt{k})$. Let P and Q be two distributions on the same support with $\mathrm{d}_{\mathrm{TV}}(P, Q) \geq \delta$. Then*

$$\mathrm{d}_{\mathrm{TV}}(P^{\otimes k}, Q^{\otimes k}) \geq \frac{\delta\sqrt{k}}{15}\,.$$

Proof of Theorem 2. Note that the sample complexity and runtime are evident from Algorithm 1. Next, note that if $\mathcal{D} = \mathcal{U}$, then $\mathcal{D}^\sigma = \mathcal{U}$ for every $\sigma \in \{\pm 1\}^n$. It now follows from Theorem 4 of [51] that `Unate-Uniformity` will output "accept" with probability at least 9/10. In particular, note that in this case, Steps 1(c) through 1(f) of `Unate-Uniformity` are identical to the `TestUniform` algorithm of [51] run on $\mathcal{U}$.

We now show how Lemma 11 implies the soundness of `Unate-Uniformity`. Suppose that the unate distribution $\mathcal{D}$ is ε-far from uniform. As before, we write

$$\mu_i \coloneqq \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}}[\boldsymbol{x}_i]\,.$$

Since $\mathcal{D}$ is unate, it follows that there exists some $\sigma^* \in \{\pm 1\}^n$ such that $\mathcal{D}^{\sigma^*}$ is monotone. In particular, we can take $\sigma_i^* = \operatorname{sign}(\mu_i)$, and so we have $\sigma_i^*\mu_i = |\mu_i| \geq 0$. Since the TV distance is preserved under bijections, $\mathcal{D}^{\sigma^*}$ is also ε-far from uniform. Applying Lemma 10 to $\mathcal{D}^{\sigma^*}$ gives

$$\|\mu\|_1 = \sum_{i=1}^n \sigma_i^*\mu_i = \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}^{\sigma^*}}\left[\sum_{i=1}^n \boldsymbol{x}_i\right] \geq \varepsilon\,.$$

Recall that we use a sample of size $m_1 = \Theta(n^{3/2}/\varepsilon^2)$ in Step 1(a) and 1(b) to compute $\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_n \in \{\pm 1\}$. Set a threshold $\theta = \frac{\varepsilon}{100n}$. For every $i \in [n]$ that satisfies $|\mu_i| \geq \theta$, we will now show that

$$\mathbf{E}\left[\sigma_i^*\boldsymbol{\sigma}_i\right] \geq \Omega\left(\frac{\varepsilon}{n}\cdot\frac{n^{3/4}}{\varepsilon}\right) = \Omega(n^{-1/4})\,, \tag{3}$$

using Lemma 12. In more detail, let P be the distribution of $\sigma_i^*\boldsymbol{x}_i$ for $\boldsymbol{x}\sim\mathcal{D}$ and Q be $-P$, i.e. $\boldsymbol{y}\sim Q$ is obtained by drawing $\boldsymbol{y}'\sim P$ and setting $\boldsymbol{y} = -\boldsymbol{y}'$. Note that

$$\Pr_{\boldsymbol{y}\sim\mathcal{P}}[\boldsymbol{y}=+1] = \frac{1+|\mu_i|}{2} \qquad \text{and} \qquad \Pr_{\boldsymbol{y}\sim\mathcal{P}}[\boldsymbol{y}=-1] = \frac{1-|\mu_i|}{2}\,,$$

with swapped probabilities for Q. It follows that $\mathrm{d_{TV}}(P,Q) = |\mu_i| \geq \frac{\varepsilon}{100n}$; note also that

$$\frac{\varepsilon}{100n} \leq \frac{1}{15\sqrt{m_1}}$$

for appropriate choice of hidden constant in m_1. In particular, applying Lemma 12 gives

$$\mathrm{d_{TV}}(P^{\otimes m_1}, Q^{\otimes m_1}) \geq \Omega\left(\frac{\varepsilon}{n}\cdot\frac{n^{3/4}}{\varepsilon}\right) = \Omega(n^{-1/4})\,.$$

In order to establish Equation (3), it suffices to show that $\mathrm{d_{TV}}(P^{\otimes m_1}, Q^{\otimes m_1}) = \Theta(1)\cdot\mathbf{E}[\sigma_i^*\boldsymbol{\sigma}_i]$. Since P and Q are Bernoulli random variables, their likelihood ratio is monotone in the number of successes (i.e., $+1$ draws), and so by the Neyman–Pearson lemma, the TV-distance maximizing set is $\{\sum_\ell \boldsymbol{y}^{(\ell)} \geq 0\}$. We may assume that m_1 is odd (to avoid handling ties), and so we get

$$\begin{aligned}
\mathrm{d_{TV}}(P^{\otimes m_1}, Q^{\otimes m_1}) &= P^{\otimes m_1}\left(\sum_{\ell=1}^{m_1}\boldsymbol{y}^{(\ell)} \geq 0\right) - Q^{\otimes m_1}\left(\sum_{\ell=1}^{m_1}\boldsymbol{y}^{(\ell)} \geq 0\right) \\
&= P^{\otimes m_1}\left(\sum_{\ell=1}^{m_1}\boldsymbol{y}^{(\ell)} \geq 0\right) - P^{\otimes m_1}\left(\sum_{\ell=1}^{m_1}\boldsymbol{y}^{(\ell)} < 0\right) \qquad (4)\\
&= \mathbf{E}[\sigma_i^*\boldsymbol{\sigma}_i]\,,
\end{aligned}$$

from which Equation (3) follows readily. Note that Equation (4) relies on the fact that $P = -Q$. Finally, note also that the same argument shows that $\mathbf{E}\left[\sigma_i^*\boldsymbol{\sigma}_i\right] \geq 0$ for every $i \in [n]$, even if $|\mu_i| < \theta$.

Let $\boldsymbol{X} \coloneqq \sum_{i=1}^{n} \mu_i \cdot \boldsymbol{\sigma}_i = \sum_{i=1}^{n} |\mu_i| \cdot (\sigma_i^* \boldsymbol{\sigma}_i)$. We have

$$\mathbf{E}[\boldsymbol{X}] \geq \sum_{i \in [n]:|\mu_i| \geq \theta} |\mu_i| \cdot \mathbf{E}[\sigma_i^* \boldsymbol{\sigma}_i] \geq \Omega(n^{-1/4}) \cdot \sum_{i \in [n]:|\mu_i| \geq \theta} |\mu_i| = \Omega(n^{-1/4} \cdot \|\mu\|_1).$$

The first step above holds since both $|\mu_i|$ and $\mathbf{E}\left[\sigma_i^* \boldsymbol{\sigma}_i\right]$ are non-negative for every $i \in [n]$. The second step applies $|\mu_i| \geq \theta \implies \mathbf{E}\left[\sigma_i^* \boldsymbol{\sigma}_i\right] \geq \Omega(n^{-1/4})$. The third step holds since our choice of $\theta = \frac{\varepsilon}{100n}$ and the fact $\|\mu\|_1 \geq \varepsilon$ together imply

$$\sum_{i \in [n]:|\mu_i| \geq \theta} |\mu_i| \geq \sum_{i=1}^{n} |\mu_i| - n \cdot \theta = \|\mu\|_1 - \frac{\varepsilon}{100} \geq \Omega(\|\mu\|_1).$$

It remains to show that

$$\mathbf{Var}[\boldsymbol{X}] \leq O(n^{-1/2} \cdot \|\mu\|_1^2).$$

By Chebyshev's inequality, the variance bound above would imply that, with probability $\Omega(1)$,

$$\boldsymbol{X} \geq \mathbf{E}[\boldsymbol{X}] - O(\sqrt{\mathbf{Var}[\boldsymbol{X}]}) \geq \Omega(n^{-1/4} \cdot \|\mu\|_1) \geq \Omega(\varepsilon / n^{1/4}),$$

where the second step applies the following observations:

- $\mathbf{E}[\boldsymbol{X}] \geq \Omega(n^{-1/4} \cdot \|\mu\|_1)$, where the hidden constant in $\Omega(\cdot)$ is lower bounded by a universal constant when $m_1 = \Theta(n^{3/2}/\varepsilon^2)$ is sufficiently large.
- We will show that $\mathbf{Var}[\boldsymbol{X}] \leq O(n^{-1/2} \cdot \|\mu\|_1^2)$, where the hidden constant in $O(\cdot)$ goes to zero as the hidden constant in $m_1 = \Theta(n^{3/2}/\varepsilon^2)$ increases.
- Therefore, for some careful choice of m_1 in the algorithm, the difference $\mathbf{E}[\boldsymbol{X}] - O(\sqrt{\mathbf{Var}[\boldsymbol{X}]})$ is positive and on the order of $\Omega(n^{-1/4} \cdot \|\mu\|_1)$.

By a Chernoff bound, we can then catch this $\Omega(\varepsilon/n^{1/4})$ bias using $\widetilde{O}\left(\frac{n}{(\varepsilon/n^{1/4})^2}\right) = \widetilde{O}(n^{3/2}/\varepsilon^2)$ additional samples in Step 1(f) of the algorithm.

By Lemma 11, we have

$$\mathbf{Var}[\boldsymbol{X}] = \sum_{i=1}^{n} \mu_i^2 \mathbf{Var}[\boldsymbol{\sigma}_i] + 2 \sum_{1 \leq i < j \leq n} \mu_i \mu_j \mathbf{Cov}[\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j] \leq \sum_{i=1}^{n} \mu_i^2 \mathbf{Var}[\boldsymbol{\sigma}_i] + O(n^{-1/2}) \cdot \|\mu\|_1^2.$$

The second term above is exactly the desired upper bound. For the first term, we note that

$$\mathbf{Var}[\boldsymbol{\sigma}_i] = e^{-\Omega(m_1 \mu_i^2)},$$

so each term $\mu_i^2 \mathbf{Var}[\sigma_i]$ is at most $O(1/m_1)$. Recalling $m_1 = \Theta(n^{3/2}/\varepsilon^2)$ and $\|\mu\|_1 \geq \varepsilon$, we conclude that

$$\sum_{i=1}^{n} \mu_i^2 \mathbf{Var}[\boldsymbol{\sigma}_i] \leq O(n/m_1) = O(\varepsilon^2/\sqrt{n}) \leq O(n^{-1/2} \cdot \|\mu\|_1^2),$$

which proves the desired variance bound $\mathbf{Var}[\boldsymbol{X}] \leq O(n^{-1/2} \cdot \|\mu\|_1^2)$ and completes the proof. ◀

3.1.4 Proof of Lemma 11

Without loss of generality, we prove the lemma when the unate distribution $\mathcal{D}$ is monotone; the more general statement then follows from the observation that flipping the i^{th} coordinate of $\mathcal{D}$ leads to a flip in the distribution of $\boldsymbol{\sigma}_i$, which preserves the magnitude of the covariances.

For $i \in [n]$, we define

$$\boldsymbol{S}_i := \sum_{\ell=1}^{m} \boldsymbol{x}_i^{(\ell)} \qquad \text{and} \qquad \mu_i := \mathop{\mathbf{E}}_{\boldsymbol{x}\sim\mathcal{D}}[\boldsymbol{x}_i]\,.$$

In particular, we have $\boldsymbol{\sigma}_i = \operatorname{sign}(\boldsymbol{S}_i) = 2\cdot\mathbf{1}\{\boldsymbol{S}_i \geq 0\} - 1$, and so

$$\mathbf{Cov}[\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j] = 4\big(\mathbf{Pr}[\boldsymbol{S}_i \geq 0, \boldsymbol{S}_j \geq 0] - \mathbf{Pr}[\boldsymbol{S}_i \geq 0]\,\mathbf{Pr}[\boldsymbol{S}_j \geq 0]\big)\,. \tag{5}$$

We will prove Lemma 11 with $i = 1$ and $j = 2$; that is, we will show that

$$\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2] = O\left(\frac{1}{\sqrt{m}}\right).$$

Define

$$p_{00} := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\mathcal{D}}\left[(\boldsymbol{x}_1, \boldsymbol{x}_2) = (-1,-1)\right], \qquad p_{10} := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\mathcal{D}}\left[(\boldsymbol{x}_1, \boldsymbol{x}_2) = (+1,-1)\right],$$
$$p_{01} := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\mathcal{D}}\left[(\boldsymbol{x}_1, \boldsymbol{x}_2) = (-1,+1)\right], \qquad p_{11} := \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\mathcal{D}}\left[(\boldsymbol{x}_1, \boldsymbol{x}_2) = (+1,+1)\right].$$

Thanks to monotonicity of $\mathcal{D}$, we have $p_{00} \leq p_{10} \leq p_{11}$ and $p_{00} \leq p_{01} \leq p_{11}$. It also follows that $\mu_1 = 1 - 2(p_{00} + p_{01}) \geq 0$ and $\mu_2 = 1 - 2(p_{00} + p_{10}) \geq 0$. The proof considers two cases depending on the value of $\mu_1 + \mu_2$:

Case 1: $\mu_1 + \mu_2 > 1/2$

We record the following easy bound on $\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]$:

$$\big|\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]\big| \leq 4\min\left\{\mathbf{Pr}\left[\boldsymbol{S}_1 < 0\right], \mathbf{Pr}\left[\boldsymbol{S}_2 < 0\right]\right\}. \tag{6}$$

Assuming Equation (6), a direct application of the Hoeffding bound gives

$$\mathbf{Pr}[\boldsymbol{S}_i < 0] \leq e^{-m\mu_i^2/2}\,, \qquad \text{and so} \qquad |\,\mathbf{Cov}(s_1, s_2)| \leq 4e^{-m/32} \ll O\left(\frac{1}{\sqrt{m}}\right),$$

where we make use of the fact that $\min\{\mu_1, \mu_2\} \geq 1/4$ by the assumption $\mu_1 + \mu_2 > 1/2$.

We now justify Equation (6). It will be convenient to write

$$A := \mathbf{Pr}[\boldsymbol{S}_1 \geq 0, \boldsymbol{S}_2 \geq 0]\,, \qquad B := \mathbf{Pr}[\boldsymbol{S}_1 \geq 0]\,, \qquad \text{and} \qquad C = \mathbf{Pr}[\boldsymbol{S}_2 \geq 0]\,.$$

Note that $\mathbf{E}[\boldsymbol{\sigma}_1] = 2B - 1$ and $\mathbf{E}[\boldsymbol{\sigma}_2] = 2C - 1$. Writing $D := 1 + A - B - C$, it is readily checked that $D = \mathbf{Pr}[\boldsymbol{S}_1 < 0, \boldsymbol{S}_2 < 0]$. It follows from Equation (5) that $\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2] = 4(A - BC)$. In particular, we have

$$\frac{1}{4}\big|\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]\big| = |A - BC|\,.$$

Now,

- If $A \geq BC$, then $A - BC \leq C(1 - B) \leq 1 - B$.
- If $A < BC$, then $BC - A \leq C - (B + C - 1) = 1 - B$, since $D \geq 0$ gives $A \geq B + C - 1$.

It follows that $|\,\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]| \leq 4(1 - B) = 4\,\mathbf{Pr}[\boldsymbol{S}_1 < 0]$. By symmetry, the same holds with B and C swapped, which yields Equation (6).

Case 2: $\mu_1 + \mu_2 \le 1/2$

In this case, we will control $\mathbf{Cov}(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ via a Gaussian approximation to the sums $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. First, let $\kappa := \mathbf{Cov}[\boldsymbol{x}_1, \boldsymbol{x}_2]$ for $\boldsymbol{x} \sim \mathcal{D}$. Note that $\kappa = p_{00} + p_{11} - p_{01} - p_{10} - \mu_1\mu_2$.

We will first control the covariance of the Gaussian approximation. Consider the standardized random vector

$$\boldsymbol{T} = (\boldsymbol{T}_1, \boldsymbol{T}_2) := \frac{1}{\sqrt{m}}(\boldsymbol{S}_1 - m\mu_1, \boldsymbol{S}_2 - m\mu_2) \qquad \text{and let} \qquad \Sigma := \begin{pmatrix} \lambda_1^2 & \kappa \\ \kappa & \lambda_2^2 \end{pmatrix}$$

where $\lambda_i^2 = 1 - \mu_i^2$. We will sometimes write $\mu = (\mu_1, \mu_2)$. Finally, let ρ be the correlation between $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$, that is,

$$\rho := \frac{\kappa}{\lambda_1\lambda_2}\,.$$

Note that by the central limit theorem, $\boldsymbol{T} \to N(0, \Sigma)$ (in distribution) as $m \to \infty$. Following Equation (5), the Gaussian approximation of the correlation is

$$\mathbf{Cov}_{\approx} := 4\Big(\mathbf{Pr}\left[\boldsymbol{Z}_1 \ge -\sqrt{m}\mu_1, \boldsymbol{Z}_2 \ge -\sqrt{m}\mu_1\right] - \mathbf{Pr}\left[\boldsymbol{Z}_1 \ge -\sqrt{m}\mu_1\right]\mathbf{Pr}\left[\boldsymbol{Z}_2 \ge -\sqrt{m}\mu_2\right]\Big).$$

Note that $\mathbf{Cov}_{\approx} = \mathbf{Cov}[\boldsymbol{\tau}_1, \boldsymbol{\tau}_2]$ where $\boldsymbol{\tau}_i = \mathrm{sign}(\boldsymbol{Z}_i - \sqrt{m}\mu_i)$. Next, we will show that $\mathbf{Cov}_{\approx} = O(m^{-1/2})$ via known estimates for computing bi-variate normal probabilities. In particular, applying Equation 3.6 of [48] gives

$$\begin{aligned}
|\,\mathbf{Cov}_{\approx}\,| &= 4\left|\frac{1}{2\pi}\int_0^{\rho}\frac{1}{\sqrt{1-z^2}}\exp\Big(-\frac{1}{2}\frac{(\sqrt{m}\frac{\mu_1}{\lambda_1})^2 + (\sqrt{m}\frac{\mu_2}{\lambda_2})^2 - 2(\sqrt{m}\frac{\mu_1}{\lambda_1})(\sqrt{m}\frac{\mu_2}{\lambda_2})z}{1-z^2}\Big)\,dz\right| \\
&\le \frac{2}{\pi}\int_0^{|\rho|}\frac{1}{\sqrt{1-z^2}}\exp\Big(-\frac{m}{2}\left(\frac{(\frac{\mu_1}{\lambda_1} - \frac{\mu_2}{\lambda_2})^2}{1-z^2} + \frac{2\frac{\mu_1}{\lambda_1}\frac{\mu_2}{\lambda_2}}{1+z}\right)\Big)\,dz \qquad (7) \\
&\le \frac{2}{\pi}\int_0^{|\rho|}\frac{1}{\sqrt{1-z^2}}\exp\Big(-\frac{m}{8}(\frac{\mu_1}{\lambda_1} + \frac{\mu_2}{\lambda_2})^2\Big)\,dz \qquad (8) \\
&\le \frac{2}{\pi}\sin^{-1}|\rho|\exp\Big(-\frac{m}{8}(\mu_1+\mu_2)^2\Big) \\
&\le |\rho|\exp\Big(-\frac{m}{8}(\mu_1+\mu_2)^2\Big)\,, \qquad (9)
\end{aligned}$$

where Equation (7) relies on the fact that the integrand is larger on the positive z side, Equation (8) uses

$$\frac{(\frac{\mu_1}{\lambda_1} - \frac{\mu_2}{\lambda_2})^2}{1-z^2} + \frac{2\frac{\mu_1}{\lambda_1}\frac{\mu_2}{\lambda_2}}{1+z} \ge (\frac{\mu_1}{\lambda_1} - \frac{\mu_2}{\lambda_2})^2 + \frac{\mu_1}{\lambda_1}\frac{\mu_2}{\lambda_2} = \frac{1}{4}(\frac{\mu_1}{\lambda_1} + \frac{\mu_2}{\lambda_2})^2 + \frac{3}{4}(\frac{\mu_1}{\lambda_1} - \frac{\mu_2}{\lambda_2})^2\,,$$

and Equation (9) relies on $\lambda_i \le 1$ and $\sin^{-1}(\theta) \le \frac{\pi\theta}{2}$ for $\theta \ge 0$.

We record the following lemma:

▶ **Lemma 13.** *For monotone distributions, if $\mu_1 + \mu_2 \le \frac{1}{2}$, then $|\rho| \le \mu_1 + \mu_2$.*

Proof. Parametrize the probabilities of the monotone distribution as

$$p_{00} = a, \quad p_{01} = a + b, \quad p_{10} = a + c, \quad p_{11} = a + d.$$

Then $\mu_1 = d + c - b$, $\mu_2 = d + b - c$, so $\mu_1 + \mu_2 = 2d$ implies $d \leq \frac{1}{4}$. Now

$$\begin{aligned}
|\rho| &= \frac{|(d-b-c) - (d+c-b)(d+b-c)|}{\sqrt{(1-\mu_1^2)(1-\mu_2^2)}} \\
&\leq \frac{2}{\sqrt{3}} \left|(d-b-c) + ((b-c)^2 - d^2)\right| && (10) \\
&\leq \frac{2}{\sqrt{3}}(d + d^2) && (11) \\
&= 2d \cdot \frac{1+d}{\sqrt{3}} \leq 2d = \mu_1 + \mu_2
\end{aligned}$$

where in Equation (10) we use that $(1-\mu_1^2)(1-\mu_2^2) \geq \frac{3}{4}$, and in Equation (11) we note that $|d - b - c| \leq d$ and $|b - c| \leq d$. ◀

Combining Equation (9) with Lemma 13, we get

$$|\operatorname{\mathbf{Cov}}_{\approx}| \leq (\mu_1 + \mu_2) \exp\left(-\frac{m}{8}(\mu_1 + \mu_2)^2\right) = O\left(\frac{1}{\sqrt{m}}\right). \tag{12}$$

Next, we will control the error in the Gaussian approximation to $\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2]$ itself by applying the multivariate Berry-Eseeen central limit theorem (cf. Theorem 8). It is readily verified using Equation (5) and the triangle inequality that

$$\frac{1}{4}\left(\mathbf{Cov}[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2] - \mathbf{Cov}_{\approx}\right) \leq \sum_{i=1}^{3} |\alpha_i - \beta_i|$$

where

$$\alpha_1 = \mathbf{Pr}[\boldsymbol{S}_1 \geq 0, \boldsymbol{S}_2 \geq 0], \quad \alpha_2 = \mathbf{Pr}[\boldsymbol{S}_1 \geq 0], \quad \alpha_3 = \mathbf{Pr}[\boldsymbol{S}_2 \geq 0],$$

$$\beta_1 = \mathbf{Pr}\left[\boldsymbol{Z}_1 \geq -\sqrt{m}\mu_1, \boldsymbol{Z}_2 \geq -\sqrt{m}\mu_2\right], \quad \beta_2 = \mathbf{Pr}\left[\boldsymbol{Z}_1 \geq -\sqrt{m}\mu_1\right], \quad \beta_3 = \mathbf{Pr}\left[\boldsymbol{Z}_2 \geq -\sqrt{m}\mu_2\right].$$

Theorem 8 combined with the fact that both $\mu_1, \mu_2 \leq 1/2$ immediately gives $|\alpha_2 - \beta_2|$, $|\alpha_3 - \beta_3| = O(m^{-1/2})$. In more detail, Theorem 8 gives

$$|\alpha_2 - \beta_2| = O\left(\frac{1}{\sqrt{m}} \cdot \frac{[|\boldsymbol{S}_1 - \mu_1|^3]}{\mathbf{E}[(\boldsymbol{S}_1 - \mu_1)^2]^{3/2}}\right) = O\left(\frac{1}{\sqrt{m}} \cdot \frac{(1+\mu_1)^3]}{(1-\mu_1^2)^{3/2}}\right) \leq O\left(\frac{1}{\sqrt{m}}\right).$$

An identical argument gives the same bound for $|\alpha_3 - \beta_3|$. Finally, in order to show that $|\alpha_1 - \beta_1| = O(m^{-1/2})$ using Theorem 8, it suffices to establish that $\mathbf{E}[\|\Sigma^{-1/2}(\boldsymbol{S} - \mu)\|_2^3] = O(1)$, where $\boldsymbol{S} = (\boldsymbol{S}_1, \boldsymbol{S}_2)$. To see this, first note that

$$\Sigma^{-1/2} = \frac{1}{2}\begin{bmatrix} \alpha & \beta \\ \beta & \alpha \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^{-1}$$

where $\alpha = \frac{1}{\sqrt{1+\rho}} + \frac{1}{\sqrt{1-\rho}}$ and $\beta = \frac{1}{\sqrt{1+\rho}} - \frac{1}{\sqrt{1-\rho}}$. Now,

$$\begin{aligned}
\|\Sigma^{-1/2}(\boldsymbol{S}-\mu)\|_2^2 &= \frac{1}{4}\left((\alpha\lambda_1^{-1}(\boldsymbol{S}_1-\mu_1)+\beta\lambda_2^{-1}(\boldsymbol{S}_2-\mu_2))^2+(\beta\lambda_1^{-1}(\boldsymbol{S}_1-\mu_1)+\alpha\lambda_2^{-1}(\boldsymbol{S}_2-\mu_2))^2\right)\\
&= \frac{1}{1-\rho^2}\left(\lambda_1^{-2}(\boldsymbol{S}_1-\mu_1)^2+\lambda_2^{-2}(\boldsymbol{S}_2-\mu_2)^2-\rho\lambda_1^{-1}\lambda_2^{-1}(\boldsymbol{S}_1-\mu_1)(\boldsymbol{S}_2-\mu_2)\right) \qquad (13)\\
&\le \frac{4}{1-\rho^2}\left(\lambda_1^{-2}+\lambda_2^{-2}+\rho\lambda_1^{-1}\lambda_2^{-1}\right) \qquad (14)\\
&= 4\cdot\frac{2-(\mu_1^2+\mu_2^2)+\kappa}{(1-\mu_1^2)(1-\mu_2^2)-\kappa^2}\\
&\le 4\cdot\frac{2-\frac{1}{2}+\frac{1}{4}}{\frac{3}{4}-\left(\frac{4}{9}\right)^2} < 16, \qquad (15)
\end{aligned}$$

where in Equation (13) we note that $\alpha^2+\beta^2=\frac{4}{1-\rho^2}$ and $2\alpha\beta=-\frac{4\rho}{1-\rho^2}$, in Equation (14) we use $|X_i-\mu_i|\le 2$, and in Equation (15) we recall $(1-\mu_1^2)(1-\mu_2^2)\ge\frac{3}{4}$ and $-\frac{4}{9}\le\kappa\le\frac{1}{4}$. The result follows immediately. ◀

3.2 Lower Bound

Turning to lower bounds for uniformity testing of unate distribution, we will establish the following:

▶ **Theorem 3.** *Let $\mathcal{A}$ be any algorithm which, given i.i.d. sample access to an unknown distribution $\mathcal{D}$, has the performance guarantee from Theorem 2 with $\varepsilon = 1-n^{-10}$. Then $\mathcal{A}$ must draw $\Omega(n^{3/2}/\log^2 n)$ samples from $\mathcal{D}$.*

In particular, Theorem 3 implies that the algorithm from Section 3.1 has essentially optimal sample complexity. The proof is deferred to the full version [45].

4 Testing Unateness with Subcube Conditioning

We now turn to the proofs of Theorems 5 and 6, starting with the former.

4.1 Upper Bound

We start by recalling Theorem 5:

▶ **Theorem 5.** *Let $\varepsilon\in(0,1)$. There is an algorithm,* `Subcube-Unate` *(Algorithm 2) which, given subcube query access to an unknown distribution $\mathcal{D}$ over $\{\pm1\}^n$, makes $\widetilde{O}(n^{3/2}/\varepsilon^2)$ queries and has the following performance guarantee:*

- *If $\mathcal{D}$ is unate, it outputs "accept" with probability at least $2/3$.*
- *If $\mathcal{D}$ is ε-far in TV distance from every unate distribution on $\{\pm1\}^n$, then it outputs "reject" with probability at least $2/3$.*

4.1.1 Preliminaries from [21]

Chakrabarty, Chen, Ristic, Seshadhri, and Waingarten [21] gave an $\widetilde{O}(n/\varepsilon^2)$-query algorithm for monotonicity testing of a distribution over $\{\pm1\}^n$ in the subcube conditional model. Our $\widetilde{O}(n^{3/2}/\varepsilon^2)$-query algorithm for unateness testing will rely on their principal technical lemma, which we state after introducing some relevant notation.

For any $x \in \{\pm 1\}^n$ and $i \in [n]$, we define

$$\delta_{x,i} \coloneqq \frac{\mathcal{D}(x^{i\to -1}) - \mathcal{D}(x^{i\to +1})}{\mathcal{D}(x^{i\to -1}) + \mathcal{D}(x^{i\to +1})} \in [-1,1]$$

as the bias on the edge $\{x, x^{\oplus i}\}$. Note that if $\mathcal{D}$ is monotone, we always have $\delta_{x,i} \le 0$, so a positive $\delta_{x,i}$ witnesses a violation of monotonicity.

Central to the analysis in [21] is the following lemma, which is implicit in the proof of [21, Lemma 2.3]. For $x \in \mathbb{R}$, we write $(x)_+ \coloneqq \max\{x, 0\}$.

▶ **Lemma 14** ([21]). *For any $\mathcal{D}$ that is ε-far from monotone, there exists a positive integer $w \le O(\log(n/\varepsilon))$ such that*

$$\mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x}\sim\mathcal{D}\\ \boldsymbol{i}\sim[n]}}\left[(\delta_{\boldsymbol{x},\boldsymbol{i}})_+^2 \ge 2^{-w}\right] \ge \widetilde{\Omega}\left(\frac{2^w\cdot\varepsilon^2}{n}\right).$$

The proof of Lemma 14 relies on a real-valued "directed" isoperimetric inequality proved in [21]. In particular, it follows from Lemma 14 that if we draw a random $\boldsymbol{w} \le O(\log(n/\varepsilon))$ with probability $\boldsymbol{w} = w$ proportional to 2^{-w}, we have

$$\mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D},\boldsymbol{i}\sim[n]}\left[(\delta_{\boldsymbol{x},\boldsymbol{i}})_+^2 \ge 2^{-\boldsymbol{w}}\right] \ge \sum_{w=1}^{O(\log(n/\varepsilon))} 2^{-w}\cdot \mathop{\mathbf{Pr}}_{\substack{\boldsymbol{x}\sim\mathcal{D}\\ \boldsymbol{i}\sim[n]}}\left[(\delta_{\boldsymbol{x},\boldsymbol{i}})_+^2 \ge 2^{-w}\right] \ge \widetilde{\Omega}\left(\frac{\varepsilon^2}{n}\right). \tag{16}$$

Give the lemma above, the algorithm of [21] is natural: we simply sample $\widetilde{O}(n/\varepsilon^2)$ triples $(\boldsymbol{w},\boldsymbol{x},\boldsymbol{i})$, in the hope of finding at least one with $(\delta_{\boldsymbol{x},\boldsymbol{i}})_+^2 \ge 2^{-\boldsymbol{w}}$. For each triple, we make $\widetilde{O}(2^w)$ queries to the subcube (or edge) $\{x, x^{\oplus i}\}$ to check whether $(\delta_{x,i})_+^2 \ge 2^{-w}$ holds. If this holds for any triple, we reject distribution $\mathcal{D}$. Over the randomness in $\boldsymbol{w}$, we make $\sum_{w=1}^{O(\log(n/\varepsilon))} 2^{-w}\cdot\widetilde{O}(2^w) = \widetilde{O}(1)$ queries for each triple, so the overall query complexity is $\widetilde{O}(n/\varepsilon^2)$.

4.1.2 Proof of Theorem 5

We first introduce some additional notation before proving Theorem 5. Recall from Notation 7 that, for $\sigma \in \{\pm 1\}^n$, $\mathcal{D}^\sigma$ is the distribution of

$$\boldsymbol{x}^\sigma \coloneqq (\sigma_1\boldsymbol{x}_1, \sigma_2\boldsymbol{x}_2, \ldots, \sigma_n\boldsymbol{x}_n),$$

where $\boldsymbol{x} \sim \mathcal{D}$, i.e., $\mathcal{D}^\sigma$ is obtained from $\mathcal{D}$ by flipping the coordinates indexed by $\{i \in [n] : \sigma_i = -1\}$. Recall that, with respect to distribution $\mathcal{D}$, we defined

$$\delta_{x,i} \coloneqq \frac{\mathcal{D}(x^{i\to -1}) - \mathcal{D}(x^{i\to +1})}{\mathcal{D}(x^{i\to -1}) + \mathcal{D}(x^{i\to +1})}.$$

In the following, we will write $\delta_{x,i}^{\mathcal{D}}$ to avoid confusion and emphasize dependence on $\mathcal{D}$. While the next identity is notation-heavy, it states an intuitive fact: if $\mathcal{D}^\sigma$ has a bias on an edge adjacent to x along the i^{th} direction, $\mathcal{D}$ should also have a bias at x^σ, albeit flipped by σ_i. More formally,

$$\begin{aligned}
\delta_{x,i}^{\mathcal{D}^\sigma} &= \frac{\mathcal{D}^\sigma(x^{i\to -1}) - \mathcal{D}^\sigma(x^{i\to +1})}{\mathcal{D}^\sigma(x^{i\to -1}) + \mathcal{D}^\sigma(x^{i\to +1})} \\
&= \frac{\mathcal{D}((x^{i\to -1})^\sigma) - \mathcal{D}((x^{i\to +1})^\sigma)}{\mathcal{D}((x^{i\to -1})^\sigma) + \mathcal{D}((x^{i\to +1})^\sigma)} && (\mathcal{D}^\sigma(x) = \mathcal{D}(x^\sigma)) \\
&= \frac{\sigma_i\cdot\left[\mathcal{D}((x^\sigma)^{i\to -1}) - \mathcal{D}((x^\sigma)^{i\to +1})\right]}{\mathcal{D}((x^\sigma)^{i\to -1}) + \mathcal{D}((x^\sigma)^{i\to +1})} \\
&= \sigma_i\cdot\delta_{x^\sigma,i}^{P}. && (17)
\end{aligned}$$

We now record a corollary of Lemma 14 which will be crucial to our analysis of Algorithm 2. Since $\mathcal{D}$ is ε-far from unate, for every sign pattern $\sigma \in \{\pm 1\}^n$, the distribution $\mathcal{D}^\sigma$ is ε-far from monotone. Applying Lemma 14 (or more precisely, Equation (16)) to $\mathcal{D}^\sigma$ then gives

$$\begin{aligned}\widetilde{\Omega}\left(\frac{\varepsilon^2}{n}\right) &\leq \mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D}^\sigma,\boldsymbol{i}\sim[n]}\left[(\delta^{\mathcal{D}^\sigma}_{\boldsymbol{x},\boldsymbol{i}})^2_+ \geq 2^{-\boldsymbol{w}}\right] \\ &= \mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D}^\sigma,\boldsymbol{i}\sim[n]}\left[(\sigma_i \cdot \delta^{\mathcal{D}}_{\boldsymbol{x}^\sigma,\boldsymbol{i}})^2_+ \geq 2^{-\boldsymbol{w}}\right] && \text{(Equation (17))} \\ &= \mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D},\boldsymbol{i}\sim[n]}\left[(\sigma_i \cdot \delta^{\mathcal{D}}_{\boldsymbol{x},\boldsymbol{i}})^2_+ \geq 2^{-\boldsymbol{w}}\right], && (18)\end{aligned}$$

where Equation (18) relies on the fact that for $\boldsymbol{x} \sim \mathcal{D}^\sigma$, the random variable $\boldsymbol{x}^\sigma$ is distributed as $\mathcal{D}$. We record this consequence of Lemma 14 for future use:

▶ **Corollary 15.** *Suppose $\mathcal{D}$ is ε-far from unate. Then for every $\sigma \in \{\pm 1\}^n$, we have*

$$\mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D},\boldsymbol{i}\sim[n]}\left[(\sigma_i \cdot \delta^{\mathcal{D}}_{\boldsymbol{x},\boldsymbol{i}})^2_+ \geq 2^{-\boldsymbol{w}}\right] \geq \widetilde{\Omega}\left(\frac{\varepsilon^2}{n}\right).$$

Corollary 15 suggests a natural test to detect distributions that are far from unate and forms the basis of Algorithm 2: iterating over each coordinate $i \in [n]$, we reject if there exist two pairs $(\boldsymbol{w}, \boldsymbol{x})$ and $(\boldsymbol{w}', \boldsymbol{x}')$ such that

$$(\delta^{\mathcal{D}}_{\boldsymbol{x},i})^2_+ \geq 2^{-\boldsymbol{w}} \quad \text{and} \quad (-\delta^{\mathcal{D}}_{\boldsymbol{x}',i})^2_+ \geq 2^{-\boldsymbol{w}'}. \tag{19}$$

Intuitively, the former condition rules out the possibility that $\mathcal{D}$ is monotone in direction i, and the latter rules out that $\mathcal{D}$ is anti-monotone in direction i. Note that if $\mathcal{D}$ is unate, then both conditions above can never hold simultaneously.

We now turn to the proof of Theorem 5:

Proof of Theorem 5. For a fixed triple $(i, \boldsymbol{x}, \boldsymbol{w})$, the expected number of conditional samples is

$$\sum_{t=1}^{\Theta(\log(n/\varepsilon))} \mathbf{Pr}[\boldsymbol{w} = t] \cdot T(t) = \sum_t \Theta(2^{-t}) \cdot \Theta(2^t \log(nk)) = \widetilde{O}(1).$$

We perform $\Theta(n \cdot k) = \widetilde{\Theta}(n^{3/2}/\varepsilon^2)$ triples, so the total query complexity is $\widetilde{O}(n^{3/2}/\varepsilon^2)$.

By Hoeffding's inequality,

$$\mathbf{Pr}\left[\left|\boldsymbol{\tau}_\ell - \delta^{\mathcal{D}}_{\boldsymbol{x},i}\right| \;\geq\; \frac{1}{8}\,2^{-\boldsymbol{w}/2}\right] \;\leq\; 2\exp\Big(-\tfrac{1}{2}\cdot(\frac{1}{8}\,2^{-\boldsymbol{w}/2})^2\,T(\boldsymbol{w})\Big) \;\leq\; \frac{\delta_{\mathrm{cmp}}}{6nk},$$

for a sufficiently large absolute constant in the definition of $T(\boldsymbol{w})$. Taking a union bound over all nk trials, with probability at least $1 - \delta_{\mathrm{cmp}}/6$ we have, simultaneously for all trials,

$$\boldsymbol{\tau}_\ell \in \Big[\delta^{\mathcal{D}}_{\boldsymbol{x},i} - \frac{1}{8}\,2^{-\boldsymbol{w}/2},\; \delta^{\mathcal{D}}_{\boldsymbol{x},i} + \frac{1}{8}\,2^{-\boldsymbol{w}/2}\Big]. \tag{20}$$

For the remainder of the argument, we condition on the event that Equation (20) holds for all trials.

Algorithm 2 The `Subcube-Unate` algorithm.

Input: Subcube query access to $\mathcal{D}$, distance parameter $\varepsilon \in (0,1)$

Output: "Accept" or "reject"

`Subcube-Unate`$(\mathcal{D}, \varepsilon)$:

1. Set $k := \widetilde{\Theta}\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ and repeat Step 2 $\Theta(1)$-times:
2. For $i \in [n]$:
   a. For $\ell \in [k]$:
      i. Sample $\boldsymbol{x} \sim \mathcal{D}$ and consider the restriction $\boldsymbol{\rho} \in \{\pm 1, *\}^n$ given by $\boldsymbol{\rho}_j = \boldsymbol{x}_j$ if $j \neq i$ and $\boldsymbol{\rho}_i = *$.
      ii. Draw $\boldsymbol{w} \in [\Theta(\log(n/\varepsilon))]$ where $\mathbf{Pr}[\boldsymbol{w} = t] \propto 2^{-t}$.
      iii. Query the conditional distribution $\mathcal{D}\,|_{\boldsymbol{\rho}}$ independently

$$T(\boldsymbol{w}) := \Theta(1) \cdot 2^{\boldsymbol{w}} \cdot \ln(12nk/\delta_{\mathrm{cmp}})$$

      times for $\delta_{\mathrm{cmp}} := n^{-3}$. Let $N_+^{(\ell)}$ and $N_-^{(\ell)}$ be the numbers of returns with coordinate i equal to $+1$ and -1, respectively ($\boldsymbol{N}_+^{(\ell)} + \boldsymbol{N}_-^{(\ell)} = T(\boldsymbol{w})$). Define the empirical edge bias

$$\boldsymbol{\tau}_\ell \;:=\; \frac{\boldsymbol{N}_-^{(\ell)} - \boldsymbol{N}_+^{(\ell)}}{\boldsymbol{N}_-^{(\ell)} + \boldsymbol{N}_+^{(\ell)}}\,, \qquad \text{and set} \;\; c_{\boldsymbol{w}} := \frac{2^{-\boldsymbol{w}/2}}{4}\,.$$

      Mark the trial ℓ as a *positive witness* if $\boldsymbol{\tau}_\ell \geq c_{\boldsymbol{w}}$, and as a *negative witness* if $\boldsymbol{\tau}_\ell < -c_{\boldsymbol{w}}$.
   b. If there exist $\ell_1, \ell_2 \in [k]$ such that ℓ_1 is a positive witness and ℓ_2 is a negative witness, then halt and output "reject."
3. If the algorithm has not rejected, output "accept."

Completeness. Conditioned on the above event, it is readily seen that `Subcube-Unate` has one-sided error. In particular, suppose $\mathcal{D}$ is a unate distribution. It follows that there exists $\sigma \in \{\pm 1\}^n$ such that $\mathcal{D}^\sigma$ is monotone. In particular, $\sigma_i \cdot \delta^{\mathcal{D}}_{x,i} \leq 0$ for all $x \in \{\pm 1\}^n$ and $i \in [n]$. Fix a coordinate $i \in [n]$ and suppose $\sigma_i = +1$. We have $\delta^{\mathcal{D}}_{\boldsymbol{x},i} \leq 0$ for every $\boldsymbol{x}$, and conditioned on Equation (20),

$$\boldsymbol{\tau}_\ell \leq \; \tfrac{1}{8}\, 2^{-\boldsymbol{w}/2} \;<\; c_{\boldsymbol{w}} = \frac{2^{-\boldsymbol{w}/2}}{4}\,,$$

so this trial cannot create a positive witness for coordinate i. An identical argument gives that if $\sigma_i = -1$, then the trial cannot create a negative witness for coordinate i. It follows that, conditioned on Equation (20), the algorithm will output "accept" with probability 1.

Soundness. We now turn to soundness of `Subcube-Unate`$(\mathcal{D}, \varepsilon)$. Suppose $\mathcal{D}$ is ε-far from every unate distribution, we will show that `Subcube-Unate`$(\mathcal{D}, \varepsilon)$ will output "reject" with probability at least 9/10. For $\boldsymbol{w}$ distributed as in Step 2(a).ii of Algorithm 2 and $i \in [n]$, let

$$q_i^+ := \mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D}}\left[(\delta^{\mathcal{D}}_{\boldsymbol{x},i})_+^2 \geq 2^{-\boldsymbol{w}}\right] \qquad \text{and} \qquad q_i^- := \mathop{\mathbf{Pr}}_{\boldsymbol{w},\boldsymbol{x}\sim\mathcal{D}}\left[(-\delta^{\mathcal{D}}_{\boldsymbol{x},i})_+^2 \geq 2^{-\boldsymbol{w}}\right]$$

be the probabilities that, over the randomness in $\boldsymbol{w}$ and $\boldsymbol{x}$, the pair $(\boldsymbol{w}, \boldsymbol{x})$ witnesses that the i^{th} coordinate cannot be monotone and anti-monotone respectively. Corollary 15 directly implies that

$$m := \frac{1}{n}\sum_{i=1}^{n} m_i \geq \widetilde{\Omega}\left(\frac{\varepsilon^2}{n}\right), \qquad \text{where } m_i := \min\left\{q_i^+, q_i^-\right\}. \tag{21}$$

Let r_i be the probability that $\texttt{Subcube-Unate}(\mathcal{D}, \varepsilon)$ outputs "reject" in Step 2(b) on coordinate i. Thanks to independence, we have

$$\mathbf{Pr}\left[\texttt{Subcube-Unate}(\mathcal{D}, \varepsilon) \text{ outputs "accept"}\right] = \prod_{i=1}^{n}(1 - r_i) \leq \exp\left(-\sum_{i=1}^{n} r_i\right),$$

and so the remainder of the argument will establish that $\sum_{i=1}^{n} r_i = \Omega(1)$.

First, note that

$$r_i \geq \left(1 - (1 - q_i^+)^{k/2}\right)\left(1 - (1 - q_i^-)^{k/2}\right) \geq \left(1 - (1 - m_i)^{k/2}\right)^2, \tag{22}$$

where the first inequality holds because the probability of returning "reject" after k iterations of Steps 2(a).i–iii is at least the probability of detecting a positive witness among the first $k/2$ iterations and a negative witness among the last $k/2$ iterations. The following lemma will be useful:

▶ **Lemma 16.** *For $K \geq 1$, the function $f(x) = (1 - (1 - x)^K)^2$ is convex on the interval $(0, \ln(2)/K)$.*

Proof. We explicitly compute the first and second derivatives of the function f. Note that

$$f'(x) = 2K\Big(1 - (1 - x)^K\Big)(1 - x)^{K-1}.$$

Taking another derivative gives

$$\begin{aligned} f''(x) &= 2K^2(1 - x)^{2K-2} - 2K(K - 1)(1 - (1 - x)^K)(1 - x)^{K-2} \\ &= 2K(1 - x)^{K-2}((2K - 1)(1 - x)^K - (K - 1)). \end{aligned}$$

So it suffices to show

$$(2K - 1)\Big(1 - \tfrac{\ln 2}{K}\Big)^K \geq K - 1,$$

or equivalently $F(K) \geq 0$ where $F(x) := x\ln\Big(1 - \frac{\ln 2}{x}\Big) - \ln\Big(\frac{x-1}{2x-1}\Big)$. Since

$$\lim_{x \to +\infty} F(x) = -\ln 2 - \ln\tfrac{1}{2} = 0,$$

showing that $F'(x) < 0$ for all $x \in (1, +\infty)$ will complete the proof.

Note that

$$F'(x) = \ln\Big(1 - \frac{1}{x/\ln 2}\Big) + \frac{1}{x/\ln 2 - 1} - \frac{1}{(x-1)(2x-1)}.$$

Furthermore, the function $h(y) := \ln\Big(1 - \frac{1}{y}\Big) + \frac{1}{y-1}$ has derivative $h'(y) = -\frac{1}{y(y-1)^2}$, and is thus strictly decreasing on $(1, +\infty)$. If we could show that

$$h(x) = \ln\Big(1 - \frac{1}{x}\Big) + \frac{1}{x-1} \leq \frac{1}{(x-1)(2x-1)} \tag{23}$$

holds for all $x \in (1, +\infty)$, we would have the desired inequality

$$F'(x) = h(x/\ln 2) - \frac{1}{(x-1)(2x-1)} < h(x) - \frac{1}{(x-1)(2x-1)} \leq 0.$$

Equation (23) follows immediately from the well-known/readily verified inequality $2a \leq (2+a) \cdot \ln(1+a)$ for $a \geq 0$ by taking $a = (x-1)^{-1}$. ◀

Returning to the soundness of `Subcube-Unate`, note that if there exists a coordinate j with $m_j \geq \frac{2\ln 2}{k}$, then thanks to Equation (22) we have

$$r_j \geq (1-(1-m_j)^{k/2})^2 \geq \left(1 - \left(1 - \frac{\ln 2}{k/2}\right)^{k/2}\right)^2 \geq (1 - e^{-\ln 2})^2 = \frac{1}{4}.$$

Consequently, the amplification in Step 1 ensures that one of the $\Theta(1)$ independent runs will output "reject" with probability at least $9/10$.

On the other hand, suppose $m_i < \frac{2\ln 2}{k}$ for all $i \in [n]$. Thanks to Lemma 16 and Jensen's inequality, we then have

$$\sum_{i=1}^{n} r_i \geq \sum_{i=1}^{n} \left(1 - (1-m_i)^{k/2}\right)^2 \geq n\left(1 - (1-m)^{k/2}\right)^2.$$

It follows that $\sum_{i=1}^{n} r_i \geq 99/100$ from Equation (21) and our choice of k by an easy application of Bernoulli's inequality $(1+u)^r \geq 1 + ru$ for $r, u \geq 0$. ◀

4.2 Lower Bound

We prove the following lower bound for testing unateness of distributions in the subcube model, which applies to all testers that only make subcube queries on either the entire hypercube (to get an unconditional sample) or a small subcube.

▶ **Theorem 6.** *There exists a constant $\varepsilon_0 > 0$ such that the following holds: If an algorithm tests whether an unknown distribution $\mathcal{D}$ over $\{\pm 1\}^n$ is unate or ε_0-far from unate by drawing Q_1 samples from $\mathcal{D}$ and making Q_2 (possibly adaptive) subcube queries on subcubes of dimensions $\leq d$, we have*

$$Q_1 + 2^d Q_2 \geq \Omega(n^{2/3}/\log^3 n).$$

Recall that our tester for Theorem 5, like the monotonicity tester of [21], only makes subcube queries on edges (i.e., one-dimensional subcubes) in addition to drawing unconditional samples. In other words, our tester satisfies the precondition of Theorem 6 with $d = 1$. The theorem then implies that all such testers must make at least $\widetilde{\Omega}(n^{2/3})$ queries in total.

The proof is deferred to the full version [45].

References

1 Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. *Advances in Neural Information Processing Systems*, 28, 2015.

2 Michał Adamaszek, Artur Czumaj, and Christian Sohler. Testing monotone continuous distributions on high-dimensional real cubes. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 56–65. SIAM, 2010.

3 Tomer Adar, Eldar Fischer, and Amit Levi. Improved bounds for high-dimensional equivalence and product testing using subcube queries. *arXiv preprint*, 2024. doi:10.48550/arXiv.2408.02347.

4 Maryam Aliakbarpour, Themis Gouleakis, John Peebles, Ronitt Rubinfeld, and Anak Yodpinyanee. Towards testing monotonicity of distributions over general posets. In *Conference on Learning Theory*, pages 34–82. PMLR, 2019. URL: http://proceedings.mlr.press/v99/aliakbarpour19a.html.

5 Roksana Baleshzar, Deeparnab Chakrabarty, Ramesh Krishnan S Pallavoor, Sofya Raskhodnikova, and C Seshadhri. Optimal unateness testers for real-valued functions: Adaptivity helps. *Theory of Computing*, 16(1):1–36, 2020. doi:10.4086/TOC.2020.V016A003.

6 Tugkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 381–390, 2004. doi:10.1145/1007352.1007414.

7 V. Bentkus. On the dependence of the Berry-Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113:385–402, 2003.

8 Ivona Bezáková, Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. Lower bounds for testing graphical models: Colorings and antiferromagnetic ising models. *Journal of Machine Learning Research*, 21(25):1–62, 2020. URL: https://jmlr.org/papers/v21/19-580.html.

9 Arnab Bhattacharyya, Eldar Fischer, Ronitt Rubinfeld, and Paul Valiant. Testing monotonicity of distributions over general partial orders. In *ICS*, pages 239–252, 2011. URL: http://conference.iiis.tsinghua.edu.cn/ICS2011/content/papers/38.html.

10 Rishiraj Bhattacharyya and Sourav Chakraborty. Property testing of joint distributions using conditional samples. *ACM Transactions on Computation Theory (TOCT)*, 10(4):1–20, 2018. doi:10.1145/3241377.

11 Hadley Black, Deeparnab Chakrabarty, and C Seshadhri. A $d^{1/2+o(1)}$ monotonicity tester for boolean functions on d-dimensional hypergrids. *SIAM Journal on Computing*, pages FOCS23–147, 2025.

12 Guy Blanc, Jane Lange, Ali Malik, and Li-Yang Tan. Lifting uniform learners via distributional decomposition. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 1755–1767, 2023. doi:10.1145/3564246.3585212.

13 Antonio Blanca, Zongchen Chen, Daniel Štefankovič, and Eric Vigoda. Complexity of high-dimensional identity testing with coordinate conditional sampling. *ACM Transactions on Algorithms*, 21(1):1–58, 2024. doi:10.1145/3686799.

14 N. Bshouty and C. Tamon. On the Fourier spectrum of monotone functions. *Journal of the ACM*, 43(4):747–770, 1996. doi:10.1145/234533.234564.

15 Clément L Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing*, pages 1–100, 2020.

16 Clément L Canonne, Xi Chen, Gautam Kamath, Amit Levi, and Erik Waingarten. Random restrictions of high dimensional distributions and uniformity testing with subcube conditioning. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 321–336. SIAM, 2021. doi:10.1137/1.9781611976465.21.

17 Clément L Canonne, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Testing bayesian networks. In *Conference on Learning Theory*, pages 370–448. PMLR, 2017. URL: http://proceedings.mlr.press/v65/canonne17a.html.

18 Clément L Canonne et al. Topics and techniques in distribution testing: A biased but representative sample. *Foundations and Trends® in Communications and Information Theory*, 19(6):1032–1198, 2022. doi:10.1561/0100000114.

19 Clément L Canonne, Elena Grigorescu, Siyao Guo, Akash Kumar, and Karl Wimmer. Testing k-monotonicity: The rise and fall of boolean functions. *Theory of Computing*, 15(1):1–55, 2019. doi:10.4086/TOC.2019.V015A001.

20 Clément L Canonne, Dana Ron, and Rocco A Servedio. Testing probability distributions using conditional samples. *SIAM Journal on Computing*, 44(3):540–616, 2015. doi:10.1137/130945508.

21 Deeparnab Chakrabarty, Xi Chen, Simeon Ristic, C. Seshadhri, and Erik Waingarten. Monotonicity testing of high-dimensional distributions with subcube conditioning. In *Symposium on Theory of Computing (STOC)*, pages 1019–1030, 2025. doi:10.1145/3717823.3718297.

22 Deeparnab Chakrabarty and C Seshadhri. A $o(n)$ monotonicity tester for boolean functions over the hypercube. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, pages 411–418, 2013. doi:10.1145/2488608.2488660.

23 Deeparnab Chakrabarty and C Seshadhri. A $O(n)$ non-adaptive tester for unateness. *arXiv preprint*, 2016. arXiv:1608.06980.

24 Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah. On the power of conditional samples in distribution testing. *SIAM Journal on Computing*, 45(4):1261–1296, 2016. doi:10.1137/140964199.

25 Mark Chen, Xi Chen, Hao Cui, William Pires, and Jonah Stockwell. Boolean function monotonicity testing requires (almost) $n^{1/2}$ queries. *arXiv preprint*, 2025. doi:10.48550/arXiv.2511.04558.

26 Xi Chen, Anindya De, Yuhao Li, Shivam Nadimpalli, and Rocco A Servedio. Mildly exponential lower bounds on tolerant testers for monotonicity, unateness, and juntas. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4321–4337. SIAM, 2024. doi:10.1137/1.9781611977912.151.

27 Xi Chen, Rajesh Jayaram, Amit Levi, and Erik Waingarten. Learning and testing junta distributions with sub cube conditioning. In *Conference on Learning Theory*, pages 1060–1113. PMLR, 2021. URL: http://proceedings.mlr.press/v134/chen21b.html.

28 Xi Chen and Cassandra Marcussen. Uniformity testing over hypergrids with subcube conditioning. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4338–4370. SIAM, 2024. doi:10.1137/1.9781611977912.152.

29 Xi Chen, Rocco A Servedio, and Li-Yang Tan. New algorithms and lower bounds for monotonicity testing. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 286–295. IEEE, 2014. doi:10.1109/FOCS.2014.38.

30 Xi Chen and Erik Waingarten. Testing unateness nearly optimally. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 547–558, 2019. doi:10.1145/3313276.3316351.

31 Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond talagrand functions: new lower bounds for testing monotonicity and unateness. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 523–536, 2017. doi:10.1145/3055399.3055461.

32 Xi Chen, Erik Waingarten, and Jinyu Xie. Boolean unateness testing with $\tilde{O}(n^{3/4})$ adaptive queries. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 868–879, 2017. doi:10.1109/FOCS.2017.85.

33 Zhihuai Chen, Siyao Guo, Qian Li, Chengyu Lin, and Xiaoming Sun. New distinguishers for negation-limited weak pseudorandom functions. *arXiv preprint*, 2022. doi:10.48550/arXiv.2203.12246.

34 Constantinos Daskalakis, Nishanth Dikkala, and Gautam Kamath. Testing ising models. *IEEE Transactions on Information Theory*, 65(11):6829–6852, 2019. doi:10.1109/TIT.2019.2932255.

35 Constantinos Daskalakis and Qinxuan Pan. Square hellinger subadditivity for bayesian networks and its applications to identity testing. In *Conference on Learning Theory*, pages 697–703. PMLR, 2017. URL: http://proceedings.mlr.press/v65/daskalakis17a.html.

36 Anindya De, Huan Li, Shivam Nadimpalli, and Rocco A Servedio. Detecting low-degree truncation. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1027–1038, 2024. doi:10.1145/3618260.3649633.

37 Anindya De, Shivam Nadimpalli, and Rocco A. Servedio. Testing convex truncation. *Mathematical Statistics and Learning*, 2025. Preliminary version in *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. doi:10.4171/MSL/50.

38 O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samordinsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000. doi:10.1007/S004930070011.

39 Elena Grigorescu, Akash Kumar, and Karl Wimmer. Flipping out with many flips: Hardness of testing k-monotonicity. *SIAM Journal on Discrete Mathematics*, 33(4):2111–2125, 2019. doi:10.1137/18M1217978.

40 William He and Shivam Nadimpalli. Testing junta truncation. *arXiv preprint*, 2023. doi:10.48550/arXiv.2308.13992.

41 Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric-type theorems. *SIAM Journal on Computing*, 47(6):2238–2276, 2018. doi:10.1137/16M1065872.

42 Subhash Khot and Igor Shinkar. An $\tilde{O}(n)$ Queries Adaptive Tester for Unateness. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)*, volume 60 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 37:1–37:7. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPIcs.APPROX-RANDOM.2016.37.

43 Jane Lange, Ronitt Rubinfeld, and Arsen Vasilyan. Properly learning monotone functions via local correction. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 75–86. IEEE, 2022. doi:10.1109/FOCS54457.2022.00015.

44 Jane Lange and Arsen Vasilyan. Agnostic proper learning of monotone functions: beyond the black-box correction barrier. *SIAM Journal on Computing*, pages FOCS23–1, 2025.

45 Daeho Lee, Shivam Nadimpalli, Mingda Qiao, and Ronitt Rubinfeld. Testing unate distributions, 2026. arXiv:2607.01573.

46 Kuldeep S Meel, Yash Pralhad Pote, and Sourav Chakraborty. On testing of samplers. *Advances in Neural Information Processing Systems*, 33:5753–5763, 2020.

47 Shyam Narayanan. On tolerant distribution testing in the conditional sampling model. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 357–373. SIAM, 2021. doi:10.1137/1.9781611976465.23.

48 Donald B Owen. Tables for computing bivariate normal probabilities. *The Annals of Mathematical Statistics*, 27(4):1075–1090, 1956.

49 Yash Pote and Kuldeep S Meel. On scalable testing of samplers. *Advances in Neural Information Processing Systems*, 35:28068–28079, 2022.

50 Mingda Qiao and Gregory Valiant. Learning Discrete Distributions from Untrusted Batches. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*, volume 94 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 47:1–47:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ITCS.2018.47.

51 Ronitt Rubinfeld and Rocco A. Servedio. Testing monotone high-dimensional distributions. *Random Struct. Algorithms*, 34(1):24–44, 2009. doi:10.1002/RSA.20247.

52 Ronitt Rubinfeld and Arsen Vasilyan. Monotone probability distributions over the boolean cube can be learned with sublinear samples. *11th Innovations in Theoretical Computer Science (ITCS 2020)*, 2020. doi:10.4230/LIPIcs.ITCS.2020.28.

53 M. Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube and Margulis' graph connectivity theorem. *GAFA*, 3(3):298–314, 1993.