Almost Optimal Distribution-Free Sample-Based Testing of k-Modality

Dana Ron
Tel Aviv University, Israel
danaron@tau.ac.il

Asaf Rosin
Tel Aviv University, Israel
asaf.rosin@gmail.com

Abstract

For an integer $k \geq 0$, a sequence $\sigma = \sigma_1, \ldots, \sigma_n$ over a fully ordered set is k-modal, if there exist indices $1 = a_0 < a_1 < \cdots < a_{k+1} = n$ such that for each i, the subsequence $\sigma_{a_i}, \ldots, \sigma_{a_{i+1}}$ is either monotonically non-decreasing or monotonically non-increasing. The property of k-modality is a natural extension of monotonicity, which has been studied extensively in the area of property testing. We study one-sided error property testing of k-modality in the distribution-free sample-based model. We prove an upper bound of[1] $O\left(\frac{\sqrt{kn}\log k}{\epsilon}\right)$ on the sample complexity, and an almost matching lower bound of $\Omega\left(\frac{\sqrt{kn}}{\epsilon}\right)$. When the underlying distribution is uniform, we obtain a completely tight bound of $\Theta\left(\sqrt{\frac{kn}{\epsilon}}\right)$, which generalizes what is known for sample-based testing of monotonicity under the uniform distribution.

2012 ACM Subject Classification Theory of computation → Streaming, sublinear and near linear time algorithms

Keywords and phrases Sample-based property testing, Distribution-free property testing, k-modality

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2020.27

Category RANDOM

Funding *Dana Ron*: Supported by the Israel Science Foundation (grant number 1146/18) and the Kadar-family award.
Asaf Rosin: This work is part of the author's MSc research.

Acknowledgements We would like to thank Ronitt Rubinfeld, since this work started as part of a project in her course, and Eden Kuperwasser, for being a great project partner. We also thank an anonymous reviewer for helpful comments.

1 Introduction

Monotonicity of functions has been studied extensively in the area of property testing [30, 35, 29, 46, 33, 32, 6, 45, 1, 37, 38, 31, 12, 16, 19, 20, 15, 24, 23, 41, 7, 21, 36, 18, 4, 26, 27, 13, 40, 42, 22, 25, 14, 43]. The different works vary in the domains and ranges they consider, as well as in the precise task studied (e.g., standard testing vs. tolerant testing and distance approximation). However, what is common to almost all of these results, is that they allow query access to the tested function, and the underlying distribution is uniform.

[1] Since our results hold for any $k \geq 0$, we should actually replace the term $\sqrt{kn}$ by $\sqrt{(k+1)n}$ and $\log k$ by $\log(k+2)$, but for the sake of readability, we refrain from doing so.

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020).
Editors: Jarosław Byrka and Raghu Meka; Article No. 27; pp. 27:1–27:19
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

In this work, we consider a natural extension of monotonicity: *k-modality*. Since the domain we study is $[n] = \{1, \ldots, n\}$, it is convenient to think of the tested object as being a sequence $\sigma = \sigma_1, \ldots, \sigma_n$, whose elements belong to a fully ordered set. A sequence σ is said to be *k-modal* if there exist indices $1 = a_0 < a_1 < \cdots < a_{k+1} = n$, such that for each i, the subsequence $\sigma_{a_i}, \ldots, \sigma_{a_{i+1}}$ is either monotonically non-decreasing or monotonically non-increasing. In other words, a sequence is k-modal if there are at most k "peaks" and "valleys" (excluding the endpoints). For example, the sequence $3, 2, 2, 1, 2, 3, 4, 2$ is 2-modal, but not 1-modal (see Figure 1). We shall assume for simplicity that σ is over $\mathbb{R}$, and use $\mathcal{M}_k$ to denote the set of all k-modal sequences over $\mathbb{R}$. Observe that a sequence is unate (i.e., either monotonically non-decreasing or monotonically non-increasing) if and only if it is 0-modal.

We study distribution-free sample-based testing of k-modality with one-sided error. Namely, the testing algorithm is given as input $k \geq 0$ and $\epsilon > 0$. For an arbitrary unknown distribution $p : [n] \to [0, 1]$, it is provided with a sample of pairs (i, σ_i), where i is selected i.i.d according to p. If σ is k-modal, then the algorithm should accept with probability 1, while if σ is ϵ-far from k-modality with respect to p, then the algorithm should reject with probability at least 2/3. A sequence σ is ϵ-far from k-modality with respect to p, if $\sum_{i:\tau_i \neq \sigma_i} p(i) > \epsilon$ for every k-modal sequence $\tau = \tau_1, \ldots, \tau_n$.

Thus, the algorithm cannot select the symbols of σ that it observes (as in the case when queries are allowed), and it must work for every underlying distribution p. Since we require that the testing algorithm have one-sided error, it may reject σ only if the sample contains evidence that σ is not k-modal. Therefore, the question we address is:

> What is the minimum sample size $s = s(n, k, \epsilon)$ such that for every sequence σ (of length n) and underlying distribution p, if σ is ϵ-far from being k-modal with respect to p, then the sample contains evidence that σ is not k-modal?

It is known that for monotonicity and unateness (i.e., the special case of $k = 0$), when the underlying distribution p is uniform, then a sample of size $\Theta(\sqrt{n/\epsilon})$ is both necessary and sufficient.[2] The question is how does the sample complexity increase as k increases, and what is the effect of having a general underlying distribution p.[3]

1.1 Our results

Our first result is the following upper bound for distribution-free testing.[4]

▶ **Theorem 1.1.** *The sample complexity of distribution-free one-sided error sample-based testing of k-modality is* $O\left(\frac{\sqrt{kn}\log k}{\epsilon}\right)$.

We show that the upper bound in Theorem 1.1 is almost tight (up to a factor of $\log k$), by establishing the next lower bound.

▶ **Theorem 1.2.** *The sample complexity of distribution-free one-sided error sample-based testing of k-modality is* $\Omega\left(\frac{\sqrt{kn}}{\epsilon}\right)$. *This lower bound holds for any $\epsilon < 1/4$ and $k \leq n/4 - 1$.*

---

[2] The lower bound for monotonicity can be found for example in [36, Claim 7.5.1] (where it actually holds for two-sided error), and this extends to unateness. The upper bound is folklore, where a more general statement, regarding any poset, follows from [33, Theorems 6&14].

[3] The problem of distribution-free sample-based testing of monotonicity is also considered in [36, Claim 7.5.2]. The upper bound they claim is correct in terms of the dependence on n (which is $\sqrt{n}$), but not in terms of the dependence on $1/\epsilon$. In our analysis, we fix the flaw in their argument.

[4] See footnote 1.

When the underlying distribution is uniform, we obtain a completely tight result, which generalizes the known result for testing monotonicity. Here, rather than having a linear dependence on $1/\epsilon$, the complexity grows like $1/\sqrt{\epsilon}$.

▶ **Theorem 1.3.** *The sample complexity of one-sided error sample-based testing of k-modality under the uniform distribution is $\Theta\left(\sqrt{\frac{kn}{\epsilon}}\right)$. The lower bounds holds for any $\epsilon < 1/4$ and $k \leq \epsilon n$.*

Note that the requirement that $k \leq \epsilon n$ is not really a constraint as we are only interested in the sublinear regime.

1.2 Techniques

The lower bounds. We start by shortly discussing our lower bounds (Theorem 1.2 for the distribution-free case, and the lower bound in Theorem 1.3, for the uniform case). They are both variants of the known lower bound for testing monotonicity with one-sided error under the uniform distribution. In both lower bounds, the sequence σ is of the form $2, 1, 4, 3, \ldots, 2m, 2m-1, 3m, \ldots, 3m$ (where the value $3m$ appears $n - 2m$ times), for an appropriate choice of m. When the underlying distribution is uniform, m is set to be $2\epsilon n$ (assuming, for simplicity, that $2\epsilon n$ is an integer). In the distribution-free case, m is set to be $n/2 - 1$, and the underlying distribution p assigns weight $\frac{2\epsilon}{m} = \frac{4\epsilon}{n-2}$ to each of the first $2m = n - 2$ symbols, and weight $\frac{1-4\epsilon}{2}$ to each of the remaining two symbols. In both cases it is not hard to verify that σ is ϵ-far from being k-modal (with respect to the corresponding distribution). What is also common to both cases is that in order to obtain evidence that σ is not k-modal, it is necessary that the sample "hit" at least $k/2$ pairs of indices $(2i-1, 2i)$ for $i \in [m]$. By a birthday-paradox-type argument, the sample must be of size $\Omega\left(\sqrt{km}/\epsilon\right)$. The lower bounds now diverge due to the difference in the setting of m, where the flexibility of the distribution-free case allows us to set m to be $\Theta(n)$ and obtain a higher lower bound.

The upper bound for the distribution-free case. We first observe that in order to test k-modality, it suffices to test two closely related properties. For the sake of simplicity, here we describe and discuss one of them, which we denote by $\mathcal{F}_t^{\Uparrow}$ where $t = k + 3$. A sequence $\sigma = \sigma_1, \ldots, \sigma_n$ belongs to $\mathcal{F}_t^{\Uparrow}$ if and only if there is no subsequence of indices $x_1, \ldots, x_t$ such that $1 \leq x_1 < \cdots < x_t \leq n$ and such that $\sigma_{x_i} < \sigma_{x_{i+1}}$ for every odd i and $\sigma_{x_i} > \sigma_{x_{i+1}}$ for every even i. We show that for any distribution p, if σ is ϵ-far with respect to p from $\mathcal{F}_t^{\Uparrow}$, then a sample of size $O(\sqrt{tn}\log t/\epsilon)$ will contain such a subsequence of indices.

A central ingredient in the proof of this sample-complexity upper bound is a structural claim. It states that if a sequence σ is ϵ-far from $\mathcal{F}_t^{\Uparrow}$, then there exist t indices $1 = a_1 < a_2 < \cdots < a_t = n$ for which the following holds. For each subsequence $\sigma_{a_i}, \ldots, \sigma_{a_{i+1}}$, if i is odd, then the subsequence is at least (ϵ/t)-far (with respect to p) from being monotonically non-decreasing, and if i is even, then it is at least (ϵ/t)-far from being monotonically non-increasing. Observe that the i^{th} subsequence and the $(i+1)^{\text{th}}$ subsequence share a common symbol. This is of importance when the common symbol has relatively large weight according to p.

The next ingredient is a claim regarding the probability of obtaining evidence, for each such subsequence, concerning its non-monotonicity (in the appropriate direction). In a certain sense we are reducing the problem of testing k-modality to testing monotonicity, where there are several subtleties to address. First, when considering the task of testing monotonicity for each of these subsequences, the fact that the underlying distribution is arbitrary, means that we need to deal both with very large probabilities and with very small probabilities, which makes the analysis more complex than in the uniform case. Second, recall that each subsequence is (ϵ/t)-far from being monotone. The "stand-alone" problem of distribution-free testing of monotonicity has sample complexity that grows linearly with the inverse of the distance to monotonicity. This seems to suggest that we get a linear dependence on t (recall that $t = k + 3$), while we claim that the dependence is $\tilde{O}(\sqrt{k})$. Therefore, the reduction to the $t - 1$ instances of testing monotonicity, should be done with care. Finally, assume that we obtain evidence of non-monotonicity (in the appropriate direction) for each subsequence (where there may be overlap between neighboring subsequences due to the common symbol). We observe that we can combine these "small pieces of evidence" to infer that σ does not belong to $\mathcal{F}_t^{\Uparrow}$.

The upper bound for the uniform case. The improvement in the sample complexity for the uniform case (as compared to the distribution-free case) has two sources. The first is that the basic task of testing monotonicity requires a smaller sample when the underlying distribution is uniform. The second is that we do not apply the structural claim described above to "break" σ into predetermined subsequences and then consider the task of testing monotonicity for each of them. Instead, the subsequences are essentially determined as part of the probabilistic analysis, together with the evidence against their monotonicity. Thus, rather than taking a union bound over events of violating monotonicity in predetermined subsequences, we lower bound the probability of sampling a sufficient number of violations by analysing an appropriate sum of geometric random variables.

Specifically, we apply the Poissonization technique (see, e.g., [48, Chapter 10]), which allows us to analyze the sample as if each pair (i, σ_i) is selected independently. We then define a process that can be viewed as traversing the sequence σ while selecting the sample "on the fly", and gathering evidence against monotonicity of subsequences. We lower bound the probability (over the selection of the sample) that the process gathers sufficient pieces of evidence before σ is fully traversed.

1.3 Related results

As noted at the start of this section, there is a plethora of works on testing monotonicity and unateness. Here we focus only on those results in which the domain is the same as ours, namely $[n]$. Unless stated explicitly otherwise, the results are for testing with queries and under the uniform distribution.

Monotonicity testing over $[n]$ was first studied by Ergun et al. [30]. They gave an algorithm whose query complexity is $O(\log n/\epsilon)$. They also showed that $\Omega(\log n)$ queries are necessary for any non-adaptive comparison-based algorithm and constant ϵ. Fischer [32] proved that this lower bound actually holds for adaptive algorithms as well.

For the special case of binary sequences, k-modality is essentially the same as k-monotonicity, which was studied in [17]. To be precise, k-monotonicity of binary sequences is equivalent to $\mathcal{F}_{k+2}^{\Uparrow}$ (where $\mathcal{F}_{k+2}^{\Uparrow}$ is as defined in Section 1.2). They show that it is possible to test k-monotonicity by performing $O(k/\epsilon)$ (non-adaptive) queries. Furthermore, they prove that any one-sided error (possibly adaptive) tester for k-monotonicity over $[n]$ must have query complexity $\Omega(k/\epsilon)$. If two-sided error is allowed, then no dependence on k is necessary, and the query complexity is $\mathrm{poly}(1/\epsilon)$ [17, 14].

The related property of k-interval functions over $[0, 1]$ was studied in [39, 3]. Each such function is defined by a partition of $[0, 1]$ into (at most) k intervals, where on each interval the value of the function is constant (either 0 or 1). Observe that if we consider a discretized version of this property where the domain is $[n]$, then it is the same as $(k-2)$-modality. Balcan et al. [3] (strengthening the result of [39]) give an upper bound of $\sqrt{k} \cdot \text{poly}(1/\epsilon)$ on the sample complexity of testing this property under the uniform distribution with two-sided error. We note that if queries are allowed, then the dependence on k can be removed [39, 3]. In addition, the result can be generalized to the distribution-free case in the active testing model [3].[5]

Other related works on testing (with queries) that generalize testing monotonicity include [8] for testing local properties of d-dimensional arrays and [44, 9, 10] for testing forbidden order patterns.

Distribution-free testing of monotonicity (with queries) was studied in [37], approximating the distance of a sequence to monotonicity was studied in [45, 2], and the problem of testing monotonicity of distributions (over totally ordered domains) was addressed in [5].

1.4 Organization

We start with some general preliminaries in Section 2. In Section 3 we present several definitions and observations relating to k-modality, which are later applied in our analysis. The upper bound for distribution-free testing is provided in Section 4, and the one for the uniform distribution in Section 5. Both lower bounds are given in Appendix C.

2 Preliminaries: sequences, distances and property testing

For an integer n, let $[n] = \{1, \ldots, n\}$, and for two integers $i \leq j$, let $[i, j] = \{i, \ldots, j\}$. For a sequence $\sigma = \sigma_1 \ldots \sigma_n$ and a subset of indices $Q \subseteq [n]$, we use $\sigma_{|Q}$ to denote the subsequence of σ corresponding to the indices in Q. A *property* of sequences $\mathcal{P}$ is simply a set of sequences. We say that σ *has property* $\mathcal{P}$, or that σ *satisfies* $\mathcal{P}$, if $\sigma \in \mathcal{P}$.

In what follows, we present several notions and notations that are defined based on a probability distribution $p : [n] \to [0, 1]$. For a set $S \subseteq [n]$, let $p(S) = \sum_{i \in S} p(S)$. Whenever p is clear from the context, we refer to $p(S)$ as the *probability weight* of S, or simply the *weight* of S.

The *distance* between σ and τ *with respect to* p, denoted $\text{dist}(\sigma, \tau, p)$, is $\sum_{i:\sigma_i \neq \tau_i} p(i)$ (or ∞ if they are not of the same length). For a property $\mathcal{P}$, the distance of σ from $\mathcal{P}$ with respect to p, denoted $\text{dist}(\sigma, \mathcal{P}, p)$, is $\min_{\tau \in \mathcal{P}}\{\text{dist}(\sigma, \tau, p)\}$. We say that σ is ϵ-far from $\mathcal{P}$ with respect to p, or (more concisely) that (σ, p) is ϵ-far from $\mathcal{P}$, if $\text{dist}(\sigma, \mathcal{P}, p) > \epsilon$. Otherwise it is ϵ-close.

▶ **Definition 2.1.** *A distribution-free sample-based testing algorithm for a property $\mathcal{P}$ of sequences is given parameters n and ϵ as well as access to samples from an unknown sequence σ of length n, generated according to an unknown distribution $p : [n] \to [0, 1]$. Namely, it receives pairs (i, σ_i) where i is distributed i.i.d. according to p. The algorithm should satisfy the following:*

[5] The notion of *active testing* is an adaptation of the notion of *active learning* to the context of property testing. The algorithm is given an unlabeled sample distributed according to the underlying distribution p and it may query the labels of part of the sample points. The two complexity measures of interest are hence the (unlabeled) sample complexity, and the query complexity (number of queries performed on the sample points). When the latter equals the former, this coincides with sample-based (distribution-free) testing (referred to as *passive testing* in [3]). However, one may aim at performing fewer queries at the cost of a larger number of unlabeled samples.

- *If $\sigma \in \mathcal{P}$, then the algorithm should accept with probability at least $2/3$.*
- *If $\mathrm{dist}(\sigma, \mathcal{P}, p) > \epsilon$, then the algorithm should reject with probability at least $2/3$.*

If p is known to be the uniform distribution over $[n]$, then we simply say that the algorithm is a sample-based testing algorithm.

If the algorithm always accepts sequences that have property $\mathcal{P}$, then we say that it has one-sided error. otherwise it has two-sided error. The sample-complexity of the algorithm is the number of samples it views (as a function of ϵ and n) when performing the aforementioned task.

In the context of testing k-modality (which is a special case of testing properties that are determined by some parameter), the algorithm is also provided with the parameter k, and its sample complexity may depend on k.

▶ **Definition 2.2.** *For positive integers n and s, and a distribution $p : [n] \to [0, 1]$, we let $I_n(s, p)$ denote the random variable corresponding to a set* [6] *consisting of s indices from $[n]$ that are selected i.i.d. according to p.*

Note that in order to obtain an upper bound s on the sample complexity of distribution-free sample-based one-sided error testing for a hereditary property $\mathcal{P}$ of sequences,[7] it suffices to show that $\Pr[\sigma_{|I_n(s,p)} \notin \mathcal{P}] \geq 2/3$, for every distribution $p : [n] \to [0, 1]$ and for every sequence σ of length n that is ϵ-far from $\mathcal{P}$ with respect to p.

It will also be useful to define an additional distance measure, which we refer to as the *deletion distance*. Let σ be a sequence of length n and $p : [n] \to [0, 1]$ a weight function (so that $\sum_{i=1}^{n} p(i)$ is not necessarily 1). For a property of sequences $\mathcal{P}$ and a subset $R \subseteq [n]$, let $\mathrm{del}(\sigma, \mathcal{P}, p, R)$ denote the minimum, taken over subsets $D \subseteq R$ such that $\sigma_{|R \setminus D} \in \mathcal{P}$, of $p(D)$ (if there is no such D, then $\mathrm{del}(\sigma, \mathcal{P}, p, R) = \infty$). If $R = [n]$, then we use the shorthand $\mathrm{del}(\sigma, \mathcal{P}, p)$ for $\mathrm{del}(\sigma, \mathcal{P}, p, [n])$.

3 Definitions and observations for k-modality

In this subsection we introduce several notions that will be used in our analysis.

For a pair of indices $x, y \in [n]$ such that $x < y$, we say that (x, y) is an *ascent* (*descent*) with respect to a sequence $\sigma = \sigma_1, \ldots, \sigma_n$, if $\sigma_x < \sigma_y$ ($\sigma_x > \sigma_y$). A pair (x, y) is an *up-pair*, denoted $\Uparrow$-pair (*down-pair*, denoted $\Downarrow$-pair) if it is an ascent (descent). We say that (x, y) and (x', y') are disjoint if $\{x, y\} \cap \{x', y'\} = \emptyset$. For $\Updownarrow \in \{\Uparrow, \Downarrow\}$, we denote $\mathrm{inv}(\Uparrow) = \Downarrow$ and $\mathrm{inv}(\Downarrow) = \Uparrow$.

The next notion will aid us in characterizing k-model sequences by being "free" of certain patterns.

▶ **Definition 3.1.** *Let σ be a sequence of length n, let t be an integer, and let $1 \leq x_1 < \cdots < x_t \leq n$. We say that $(x_1, \ldots, x_t)$ is a t-upward (t-downward) subsequence with respect to σ if for every odd $i \in [t-1]$, (x_i, x_{i+1}) is an ascent (descent), and for every even $i \in [t-1]$, (x_i, x_{i+1}) is a descent (ascent).*
We shall use the symbolic shorthand t-$\Uparrow$ for t-upward sequences and t-$\Downarrow$ for t-downward sequences.

[6] Note that while the sampled indices i (as defined in Definition 2.1) are i.i.d. and hence may appear with repetitions, $I_n(s)$ is a set and hence does not include repetitions

[7] A property of sequences is hereditary if it is preserved under restrictions to subsequences.

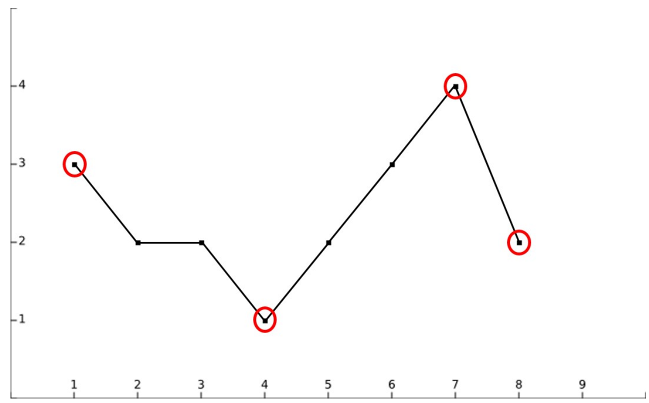

Figure 1 An illustration of the sequence $\sigma = 3, 2, 2, 1, 2, 3, 4, 2$. Notice that the marked subsequence $(1, 4, 7, 8)$ is a 4-downward (4-$\Downarrow$) subsequence with respect to σ, and that $\sigma \in \mathcal{F}_5^{\Downarrow}$.

▶ **Definition 3.2.** *For $\Updownarrow \in \{\Uparrow, \Downarrow\}$, we denote by $\mathcal{F}_t^{\Updownarrow}$ the set of all sequences σ such that there is no t-$\Updownarrow$ subsequence with respect to σ (so that $\mathcal{F}$ stands for "free"). For the special case of $t = 2$, we say that a sequence is $\Updownarrow$-monotone if it belongs to $\mathcal{F}_2^{\mathrm{inv}(\Updownarrow)}$.*

For example, if $\sigma = 3, 2, 2, 1, 2, 3, 4, 2$, then $(1, 4, 7, 8)$ is a 4-downward (4-$\Downarrow$) subsequence with respect to σ, and $\sigma \in \mathcal{F}_5^{\Downarrow}$. For an Illustration, see Figure 1.

Observe that a sequence σ is k-modal if and only if there is no $(k+3)$-$\Uparrow$ subsequence nor any $(k+3)$-$\Downarrow$ subsequence with respect to σ:

▶ **Observation 3.3.** *For any non-negative integer k we have that $\mathcal{M}_k = \mathcal{F}_{k+3}^{\Uparrow} \bigcap \mathcal{F}_{k+3}^{\Downarrow}$.*

The next notion and observation will be useful when analyzing the evidence found in a sample, that a sequence is not k-modal.

▶ **Definition 3.4.** *Let σ be a sequence of length n, let t be a positive integer, and let $1 \leq x_1 < y_1 \leq x_2 < y_2 \leq \cdots \leq x_t < y_t$. We say that $((x_1, y_1), \ldots, (x_t, y_t))$ is a t-upward-pair (t-downward-pair) sequence with respect to σ, if for every odd $i \in [t]$, (x_i, y_i) is an ascent (descent), and for every even $i \in [t]$, (x_i, y_i) is a descent (ascent). Here too we use the symbolic shorthands t-$\Uparrow$-pair (for t-upward-pair sequences) and t-$\Downarrow$-pair (for t-downward-pair sequences).*

▶ **Observation 3.5.** *For any sequence σ, integer $t \geq 2$, and $\Updownarrow \in \{\Uparrow, \Downarrow\}$, if there is a $(t-1)$-$\Updownarrow$-pair sequence with respect to σ, then there is a t-$\Updownarrow$ subsequence with respect to σ.*

To verify the validity of the last observation, consider for simplicity the case that $\Updownarrow = \Uparrow$. Given a $(t-1)$-$\Uparrow$-pair sequence $((x_1, y_1), \ldots, (x_{t-1}, y_{t-1}))$ we define a t-$\Uparrow$ subsequence $(x'_1, \ldots, x'_t)$ as follows: $x'_1 = x_1$, $x'_t = y_{t-1}$, $x'_i = \max(y_{i-1}, x_i)$ for each even $i \in [2, t-1]$, and $x'_i = \min(y_{i-1}, x_i)$ for each odd $i \in [2, t-1]$.

We shall also make use of the following observation, which will allow us to work with the deletion distance. Its simple proof is given in Appendix B.

▶ **Observation 3.6.** *Let σ be a sequence of length n, $p : [n] \to [0, 1]$ a probability distribution, and k a non-negative integer. Then $\mathrm{dist}(\sigma, \mathcal{M}_k, p) = \mathrm{del}(\sigma, \mathcal{M}_k, p)$.*

The last observation in this section will allow us to reduce the problem of testing $\mathcal{M}_k$ to testing $\mathcal{F}_{k+3}^{\Uparrow}$ and $\mathcal{F}_{k+3}^{\Downarrow}$.

▶ **Observation 3.7.** *Let σ be a sequence of length n, $p : [n] \to [0,1]$ a probability distribution, and k a non-negative integer. Then* $\mathrm{dist}(\sigma, \mathcal{M}_k, p) \leq \mathrm{del}(\sigma, \mathcal{F}^{\Uparrow}_{k+3}, p) + \mathrm{del}(\sigma, \mathcal{F}^{\Downarrow}_{k+3}, p)$.

To verify Observation 3.7, let $t = k+3$ and for each $\Updownarrow \in \{\Uparrow, \Downarrow\}$ let $D^{\Updownarrow} \subseteq [n]$ be a subset satisfying $\sigma_{|[n]\setminus D^{\Updownarrow}} \in \mathcal{F}^{\Updownarrow}_t$ and $p(D^{\Updownarrow}) = \mathrm{del}(\sigma, \mathcal{F}^{\Updownarrow}_t, p)$. Since $\mathcal{F}^{\Uparrow}_t$ and $\mathcal{F}^{\Downarrow}_t$ are hereditary properties, for $D = D^{\Uparrow} \cup D^{\Downarrow}$, we have that $\sigma_{|[n]\setminus D} \in \mathcal{F}^{\Uparrow}_t \cap \mathcal{F}^{\Downarrow}_t$. By Observation 3.3, this means that $\sigma_{|[n]\setminus D} \in \mathcal{M}_k$, so that $\mathrm{del}(\sigma, \mathcal{M}_k, p) \leq p(D^{\Updownarrow}) \leq p(D^{\Uparrow}) + p(D^{\Downarrow}) = \mathrm{del}(\sigma, \mathcal{F}^{\Uparrow}_t, p) + \mathrm{del}(\sigma, \mathcal{F}^{\Downarrow}_t, p)$. By Observation 3.6, $\mathrm{dist}(\sigma, \mathcal{M}_k, p) = \mathrm{del}(\sigma, \mathcal{M}_k, p)$ and Observation 3.7 is verified.

4 The upper bound for distribution-free testing

In this section we prove Theorem 1.1, which is restated next.

▶ **Theorem 1.1.** *The sample complexity of distribution-free one-sided error sample-based testing of k-modality is* $O\left(\frac{\sqrt{kn}\log k}{\epsilon}\right)$.

4.1 Structural claims

We start with two structural claims, where the second builds on the first (and where the first will also serve us for the upper bound under the uniform distribution).

▷ Claim 4.1. Let $\sigma = \sigma_1, \ldots, \sigma_n$ be a sequence of length n and $p : [n] \to [0,1]$ a weight function. Then for any $m \in [n]$, $t > 2$ and $\Updownarrow \in \{\Uparrow, \Downarrow\}$,

$$\mathrm{del}(\sigma, \mathcal{F}^{\Updownarrow}_t, p) \leq \mathrm{del}(\sigma, \mathcal{F}^{\Updownarrow}_2, p, [m]) + \mathrm{del}(\sigma, \mathcal{F}^{\mathrm{inv}(\Updownarrow)}_{t-1}, p, [m+1, n]) \ .$$

In order to prove this claim, we essentially show that if it is possible to partition a sequence into two parts, such that the first is free of upward pairs, and the second is free of $(t-1)$-downward sequences, then the entire sequence is free of t-upward sequences.

Proof. By the definition of the deletion distance, there exists a set $I_1 \subseteq [m]$ of weight $\mathrm{del}(\sigma, \mathcal{F}^{\Updownarrow}_2, p, [m])$, such that $\sigma_{|[m]\setminus I_1} \in \mathcal{F}^{\Updownarrow}_2$. Similarly, there exists a set $I_2 \subseteq [m+1, n]$ of weight $\mathrm{del}(\sigma, \mathcal{F}^{\mathrm{inv}(\Updownarrow)}_{t-1}, p, [m+1, n])$, such that $\sigma_{|[m+1,n]\setminus I_2} \in \mathcal{F}^{\mathrm{inv}(\Updownarrow)}_{t-1}$.

Let $\tilde{\sigma}$ denote the sequence that is obtained from σ by deleting all the indices in $I_1 \cup I_2$. Namely, $\tilde{\sigma} = \sigma_{[n]\setminus I_1 \cup I_2}$. We next show that $\tilde{\sigma} \in \mathcal{F}^{\Updownarrow}_t$. We represent $\tilde{\sigma}$ as a concatenation of two sequences, τ_1 and τ_2, where $\tau_1 = \sigma_{|[m]\setminus I_1}$ and $\tau_2 = \sigma_{|[m+1,n]\setminus I_2}$. Assume, contrary to the claim, that there exists a t-$\Updownarrow$ subsequence $(x_1, x_2 \ldots, x_t)$ with respect to $\tilde{\sigma}$. Since (x_1, x_2) is a $\Updownarrow$-pair (by the definition of a t-$\Updownarrow$ subsequence) while $\tau_1 \in \mathcal{F}^{\Updownarrow}_2$, necessarily $x_2 > |\tau_1|$. But then $(x_2, \ldots, x_t)$ is a $(t-1)$-inv($\Updownarrow$) subsequence with respect to τ_2 (more precisely, $(x_2 - |\tau_1|, \ldots, x_t - |\tau_1|)$ is such a subsequence), in contradiction to $\tau_2 \in \mathcal{F}^{\Updownarrow}_{t-1}$.

We conclude that

$$\mathrm{del}(\sigma, \mathcal{F}^{\Updownarrow}_t, p) \leq p(I_1 \cup I_2) = \mathrm{del}(\sigma, \mathcal{F}^{\Updownarrow}_2, p, [m]) + \mathrm{del}(\sigma, \mathcal{F}^{\mathrm{inv}(\Updownarrow)}_{t-1}, p, [m+1, n]) \ ,$$

as claimed. ◁

We next recursively apply Claim 4.1 to show that if σ is far from $\mathcal{F}^{\Updownarrow}_t$, then we can define $t-1$ (almost disjoint) consecutive subsequences, such that each is relatively far from either $\mathcal{F}^{\Updownarrow}_2$ or $\mathcal{F}^{\mathrm{inv}(\Updownarrow)}_2$.

▷ Claim 4.2. Let σ be a sequence of length n, $p : [n] \to [0,1]$ a weight function, $t \geq 2$, and $\Updownarrow \in \{\Uparrow, \Downarrow\}$. Denote $\Delta = \mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p)$ and suppose that $\Delta > 0$. Then there exist indices $1 = a_1 < \cdots < a_t = n$ such that $\mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, [a_i, a_{i+1}]) \geq \Delta/(t-1)$ for every odd $i \in [t-1]$, and $\mathrm{del}(\sigma, \mathcal{F}_2^{\mathrm{inv}(\Updownarrow)}, p, [a_i, a_{i+1}]) \geq \Delta/(t-1)$ for every even $i \in [t-1]$.

Note that partitioning the domain into disjoint intervals that obey the condition stated in Claim 4.2 may not be possible, due to the existence of indices with a large weight according to p. To address this issue we allow each pair of consecutive intervals ($[a_i, a_{i+1}]$ and $[a_i, a_{i+1}]$) to share a common index. This suffices for our purposes as we shall see in Section 4.3

Proof. We prove the claim by induction on t. For the base case, $t = 2$, the claim is trivial. Turning to the induction step, we assume that the claim holds for $t-1$ (where $t > 2$), and prove it for t.

Define a_2 to be the smallest index in $[n]$ satisfying the required condition $\mathrm{del}(\sigma,\mathcal{F}_2^{\Updownarrow},p,[a_2]) \geq \Delta/(t-1)$ (recall that Δ denotes $\mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p)$). Note that: (1) such an index exists, as $\mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, [n]) \geq \mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p, [n]) = \Delta \geq \Delta/(t-1)$, and (2) $a_2 > 1$, as $\sigma_1 \in \mathcal{F}_2^{\Updownarrow}$ so that $\mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, \{1\}) = 0$.

Thus, we can apply Claim 4.1 with $m = a_2 - 1$ to obtain that $\mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p) \leq \mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, [a_2 - 1]) + \mathrm{del}(\sigma, \mathcal{F}_{t-1}^{\mathrm{inv}(\Updownarrow)}, p, [a_2, n])$. By the definition of a_2, we have that $\mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, [a_2 - 1]) < \Delta/(t-1)$. We infer that $\mathrm{del}(\sigma, \mathcal{F}_{t-1}^{\mathrm{inv}(\Updownarrow)}, p, [a_2, n]) > \Delta - \Delta/(t-1) = (t-2)\Delta/(t-1)$.

The claim is established because the existence of $a_3, \ldots, a_t$ (recall that $t > 2$) satisfying the required conditions is implied by the induction hypothesis, using $\tilde{t} = t-1$, $\tilde{\Updownarrow} = \mathrm{inv}(\Updownarrow)$, $\widetilde{\sigma} = \sigma_{a_2} \ldots \sigma_n$ and $\widetilde{\Delta} = (t-2)\Delta/(t-1)$. Notice that the conditions are satisfied, as $\widetilde{\Delta}/(t-2) = \Delta/(t-1)$. ◁

4.2 Probabilistic claims – obtaining evidence of non-monotonicity

In the following claim we give conditions under which a sample contains evidence that a subsequence is not in $\mathcal{F}_2^{\Updownarrow}$ (with probability at least 2/3). We later apply this claim to the subsequences defined in Claim 4.2.

▷ Claim 4.3. Let σ be a sequence of length n, $p : [n] \to [0,1]$ a probability distribution, and R a subset of $[n]$. Suppose that for $\Updownarrow \in \{\Uparrow, \Downarrow\}$ and for positive β and δ, we have that $\mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, R) \geq \beta$, and that $p(i) \geq \delta$ for each $i \in R$. Then for $s = \Theta(1/\sqrt{\delta \cdot \beta})$, the probability over the choice of $Q = I_n(s,p)$ that $\sigma_{|R \cap Q} \notin \mathcal{F}_2^{\Updownarrow}$, is at least 2/3.

In order to prove Claim 4.3, we lower bound the probability weight of a sample that falls into a prespecified subset of the domain $[n]$.

▷ Claim 4.4. Let $p : [n] \to [0,1]$ be a probability distribution, and C a subset of $[n]$. Suppose that for positive β and $\delta \leq \beta/c$, where c is a sufficiently large constant, $p(C) \geq \beta$ and $p(x) \geq \delta$ for each $x \in C$. Then for $s = 1/\sqrt{\delta \cdot \beta}$, letting $Q = I_s(n,p)$,

$$\Pr_Q \left[p(C \cap Q) \geq \frac{\delta \beta s}{4} \right] \geq \frac{9}{10} .$$

Note that if Q and $C \cap Q$ were defined as multisets rather than sets, i.e., if we were to take repetitions into account, then Claim 4.4 would have followed from a standard tail bound (Fact A.3). However, since in our case $C \cap Q$ is a set, we need to analyse the effect of collisions in the sample. As the distribution p may contain large probabilities, the collisions can have a significant impact. In order to overcome this difficulty, we use a "flattening technique" that is similar to the one introduced in [28] (see also [34]).

Proof. We prove the claim in two stages. In the first stage we define a "flattened" probability distribution $\hat{p}$ over $[\hat{n}]$ for $\hat{n} \geq n$ together with a subset $\widehat{C} \subseteq [\hat{n}]$, for which we show the following: for any positive integer s, letting $\widehat{Q} = I_s(\hat{n}, \hat{p})$, we have that

$$\text{for any } \tau \geq 0, \quad \Pr_Q[p(C \cap Q) \geq \tau] \geq \Pr_{\widehat{Q}}[\hat{p}(\widehat{C} \cap \widehat{Q}) \geq \tau] \; . \tag{1}$$

In the second stage we show, using certain properties of $\hat{p}$ and $\widehat{C}$, that for s as in the premise of the claim,

$$\Pr_{\widehat{Q}}\left[\hat{p}(\widehat{C} \cap \widehat{Q}) \geq \frac{\delta\beta s}{4}\right] \geq \frac{9}{10} \; . \tag{2}$$

Stage I. If $p(x) \in [\delta, 2\delta]$ for each $x \in C$, then we simply let $\hat{n} = n$, $\hat{p} = p$, and $\widehat{C} = C$, so that Equation (1) holds trivially. Otherwise, the high level idea is that we "split" each $x \in C$ into a subset of indices, each with probability weight in $[\delta, 2\delta]$ (according to $\hat{p}$), as explained precisely next. Assume, without loss of generality, that $C = \{m+1, \ldots, n\}$ for some $m < n-1$. For each $x \in C$ let $\alpha(x) = \lceil \frac{p(x)}{2\delta} \rceil$. Set $\hat{n} = m + \sum_{x=m+1}^{n} \alpha(x)$ and $\widehat{C} = \{m+1, \ldots, \hat{n}\}$. It remains to define $\hat{p}$. With each $x \in C$ we associate a (disjoint) subset, denoted $J(x)$, of $\alpha(x)$ indices in $\{m+1, \ldots, \hat{n}\}$. In particular, we can let $J(x) = \{m + 1 + \sum_{x'=m+1}^{x-1} \alpha(x'), \ldots, m + \sum_{x'=m+1}^{x} \alpha(x')\}$ for each $x \in \{m+1, \ldots, n\}$. It will be useful to define a mapping $\phi : [\hat{n}] \to [n]$, where for each $y \in \widehat{C}$, $\phi(y) = x$ where x satisfies $y \in J(x)$. For $y \notin \widehat{C}$, we let $\phi(y) = y$. For each $y \notin C$ we set $\hat{p}(y) = p(y)$, and for each $y \in \widehat{C} = \{m+1, \ldots, \hat{n}\}$ we set $\hat{p}(y) = p(\phi(y))/\alpha(\phi(y))$. Note that by the definition of $\alpha(\cdot)$, this ensures that $\hat{p}(y) \in [\delta, 2\delta]$ and that $p(x) = \hat{p}(J(x))$ for each $x \in C$.

In order to establish Equation (1), we apply a coupling argument. Specifically, we define a random variable $\widetilde{Q} \subseteq [n]$ based on $\widehat{Q} \subseteq [\hat{n}]$ as follows: $\widetilde{Q} = \{\phi(y) : y \in \widehat{Q}\}$. Since for each $x \in [n]$ we have that $p(x) = \sum_{y:\phi(y)=x} \hat{p}(y)$, by its definition, $\widetilde{Q}$ is identically distributed to Q. Next observe that for each $x \in C$, if $x \notin \widetilde{Q}$, then necessarily $J(x) \cap \widehat{Q} = \emptyset$, while if $x \in \widetilde{Q}$, then $p(x) \geq \hat{p}(J(x) \cap \widehat{Q})$. Therefore, $p(C \cap \widetilde{Q}) \geq \hat{p}(\widehat{C} \cap \widehat{Q})$, and Equation (1) follows.

Stage II. Since $\hat{p}(\widehat{C}) = p(C)$, and $p(C) \geq \beta$, we have that $\hat{p}(\widehat{C}) \geq \beta$. For $s = 1/\sqrt{\delta\beta}$, let $Y_1, \ldots, Y_s$ be independent random variables such that for each $y \in [\hat{n}]$, $\Pr[Y_r = y] = \hat{p}(y)$. Let $\hat{s} = |\{r : Y_r \in \widehat{C} \setminus \{Y_1, \ldots, Y_{r-1}\}\}|$, and observe that $\hat{p}(\widehat{C} \cap \widehat{Q}) \geq \hat{s} \cdot \delta$ (since $\hat{p}(y) \geq \delta$ for each $y \in \widehat{C}$). Hence, in order to establish Equation (2), it suffices to upper bound the probability that $\hat{s} < \beta s/4$.

To this end we also define $s' = |\{r : Y_r \in \widehat{C}\}|$, so $\mathbb{E}[s'] = \hat{p}(\widehat{C}) \cdot s$. We may think of $\hat{s}$ as being determined by first determining s', and then taking s' samples from $\widehat{C}$, where each $y \in \widehat{C}$ is selected independently with probability $\hat{p}(y)/\hat{p}(\widehat{C})$. Since $s' \sim \text{Bin}(s, \hat{p}(\widehat{C}))$, we have (by Fact A.3) that $\Pr[s' < \mathbb{E}[s']/2] \leq e^{-\hat{p}(\widehat{C})s/8}$. By our setting of s and since $\delta \leq \beta/c$, this probability is at most $1/20$ for a sufficiently large constant c. We henceforth condition on the event that $s' \geq \mathbb{E}[s']/2 \geq \beta s/2$. If all Y_r that belong to C were distinct, then we would have that $\hat{s} = s'$, and we would be done. Since this is not necessarily the case, it remains to show that $\hat{s}$ is not much smaller than s'. To be precise, since our goal is to lower bound the probability that $\hat{s} < \beta s/4$, we condition on the event that $s' = \beta s/2$ (as the probability that $\hat{s} < \beta s/4$ can only decrease as s' increases).

Let $q = |\{r : Y_r \in \widehat{C} \cap \{Y_1, \ldots, Y_{r-1}\}\}|$, so that $\hat{s} = s' - q$. Observe that $\mathbb{E}[q] \leq \binom{s'}{2} \cdot \frac{2\delta}{\hat{p}(\widehat{C})} \leq s' \cdot \frac{\delta s'}{\hat{p}(\widehat{C})}$, which by our condition on s' and since $\hat{p}(\widehat{C}) \geq \beta$ is at most $\frac{\beta s}{2} \cdot \frac{\delta s}{2}$. Once again by our setting of s and since $\delta \leq \beta/c$, this is at most $\frac{\beta s}{80}$ for a sufficiently large constant c. By Markov's inequality, $\Pr\left[q > \frac{\beta s}{4}\right] \leq 1/20$. We can conclude that with probability at least $1 - 2/20 = 9/10$, $\hat{s} \geq \beta s/4$, and the claim follows. ◁

We are now ready to prove Claim 4.3, which applies Claim 4.4 to a vertex cover C in the "violation graph", as defined next.

Proof of Claim 4.3. For indices $x, y \in R$, we say that (x, y) is a "violating pair" if $x < y$ and $\sigma_x \sigma_y \notin \mathcal{F}_2^{\Updownarrow}$. Consider the "violation graph" $G_R = ([n], E_R)$, where E_R is the set of all violating pairs, and recall that a *vertex cover* of a graph is a subset C of vertices such that each edge of the graph contains at least one vertex in C. For a subset of vertices S, let $\Gamma_R(S)$ denote the set of neighbors of vertices belonging to S in the graph G_R.

Let C be a minimum-weight vertex cover of G_R with respect to p. Note that $\sigma_{|R\setminus C} \in \mathcal{F}_2^{\Updownarrow}$, so that $p(C) = \mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, R) \geq \beta$. Let $s = 16/\sqrt{\delta \cdot \beta}$. First, consider a sample, denoted Q_1, of $s/16$ vertices (indices in $[n]$) drawn independently according to p. By Claim 4.4, the probability that $p(C \cap Q_1) \geq \beta\delta s/64$ is at least $9/10$. Conditioned on this event occurring, we consider a second sample, denoted Q_2, of $15s/16$ vertices drawn independently according to p. As observed in [36], $p(\Gamma_R(C \cap Q_1)) \geq p(C \cap Q_1)$ (since otherwise, $C' = (C \setminus Q_1) \cup \Gamma_R(C \cap Q_1)$ is a vertex cover with smaller weight than C). Hence, each of these $15s/16$ sampled vertices belongs to $\Gamma_R(C \cap Q_1)$ with probability at least $p(C \cap Q_1)$. Therefore, the probability that $Q_1 \times Q_2$ contains no violating pair is upper bounded by

$$(1 - p(C \cap Q_1))^{15s/16} \leq \left(1 - \frac{\beta\delta s}{64}\right)^{15s/16} \leq e^{-15\beta\delta s^2/1024} \leq e^{-3.75} \leq 1/10 .$$

We conclude that the sample Q contains a violating pair, and hence $\sigma_{|R\cap Q} \notin \mathcal{F}_2^{\Updownarrow}$, with probability at least $1 - 1/10 - 1/10 > 2/3$, as claimed. ◁

4.3 Wrapping things up

By combining Claims 4.2 and Claim 4.3, we can establish the next lemma, which gives an upper bound on the sample complexity of one-sided error distribution-free testing of $\mathcal{F}_t^{\Updownarrow}$.

▶ **Lemma 4.5.** *Let σ be a sequence of length n, $p : [n] \to [0, 1]$ a probability distribution, $t \geq 2$, $\epsilon > 0$ and $\Updownarrow \in \{\Uparrow, \Downarrow\}$. If $\mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p) > \epsilon$, then for $s = \Theta(\sqrt{tn} \log t/\epsilon)$ and $Q = I_n(s, p)$ we have that $\Pr\left[\sigma_{|R\cap Q} \notin \mathcal{F}_t^{\Updownarrow}\right] \geq \frac{2}{3}$.*

Proof. Let $\delta = \frac{\epsilon}{2n}$ and let $B = \{x \in [n] : p(x) \geq \delta\}$. Since $p([n] \setminus B) < n \cdot \delta = \epsilon/2$, we have that $\mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p, B) > \epsilon/2$. We would like to apply Claim 4.2 to $\sigma_{|B}$, and hence we need to first define a corresponding weight function (over $[|B|]$), which we denote by $p_{|B}$. Specifically, denoting the elements in B by $\{b_1, \ldots, b_{|B|}\}$ where $b_1 < \cdots < b_{|B|}$, we let $p_{|B}(j) = p(b_j)$. Observe that $\mathrm{del}(\sigma_{|B}, \mathcal{F}_t^{\Updownarrow}, p_{|B}) = \mathrm{del}(\sigma, \mathcal{F}_t^{\Updownarrow}, p, B)$.

We can now apply Claim 4.2 to $\sigma_{|B}$ and $p_{|B}$ (as well as t and $\Updownarrow$), and obtain $1 \leq a_1 < \cdots < a_t \leq |B|$ as stated in the claim. For each $i \in [t-1]$, let $R_i = B \cap [b_{a_i}, b_{a_{i+1}}]$. Therefore, $\mathrm{del}(\sigma, \mathcal{F}_2^{\Updownarrow}, p, R_i) > \epsilon/(2(t-1))$ for each odd $i \in [t-1]$, and $\mathrm{del}(\sigma, \mathcal{F}_2^{\mathrm{inv}(\Updownarrow)}, p, R_i) > \epsilon/(2(t-1))$ for each even $i \in [t-1]$.

Consider any fixed choice of $i \in [t-1]$. In particular, assume first that i is odd. Suppose we apply Claim 4.3 with $R = R_i$, $\beta = \epsilon/(2(t-1))$, $\delta = \epsilon/(2n)$ (and $\Updownarrow$). Observe that $1/\sqrt{\delta\beta} = \Theta(\sqrt{tn}/\epsilon)$. Since in the current lemma $s = \Theta(\sqrt{tn} \log t/\epsilon)$, we get that for $Q = I_n(s, p)$ (i.e., a sample of size $c \log t$ times larger than the sample in the statement of Claim 4.3), $\Pr_Q[\sigma_{|R_i \cap Q} \in \mathcal{F}_2^{\Updownarrow}] \leq (1/3)^{c \log t} < 1/(3t)$ (for an appropriate constant c). That is, $\sigma_{|R_i \cap Q}$ does not contain an $\mathrm{inv}(\Updownarrow)$-pair with probability at most $1/(3t)$. An analogous statement holds for each even $i \in [t-1]$ (with respect to $\mathcal{F}_2^{\mathrm{inv}(\Updownarrow)}$).

By applying a union bound over all $i \in [t-1]$, we get that with probability at least $2/3$, there are $(t-1)$ pairs $(x_1, y_1), \ldots, (x_{t-1}, y_{t-1})$ such that $x_i, y_i \in R_i \cap Q$, for which the following holds: For every odd $i \in [t-1]$, the pair (x_i, y_i) is a $\Updownarrow$-pair, and for every even $i \in [t-1]$, the pair (x_i, y_i) is an inv($\Updownarrow$)-pair. Notice that for each $i \in [t-1]$, $y_i \in R_i$ and $x_{i+1} \in R_{i+1}$, so we have that $y_i \leq x_{i+1}$. Therefore (recalling Definition 3.5), with probability at least $2/3$, the sample Q contains a $(t-1)$-$\Updownarrow$-pair sequence with respect to σ. By Observation 3.5, this implies that the sample contains a t-$\Updownarrow$ subsequence with respect to σ, that is, $\sigma_{|Q} \notin \mathcal{F}_t^{\Updownarrow}$. ◀

We are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. If $\text{dist}(\sigma, \mathcal{M}_k, p) > \epsilon$, then, by Observation 3.7, for some $\Updownarrow \in \{\Uparrow, \Downarrow\}$, $\text{del}(\sigma, \mathcal{F}_{k+3}^{\Updownarrow}, p) > \epsilon/2$ (otherwise, $\text{del}(\sigma, \mathcal{F}_{k+3}^{\Uparrow}, p) + \text{del}(\sigma, \mathcal{F}_{k+3}^{\Downarrow}, p) \leq \epsilon$, so that by Observation 3.7, $\text{dist}(\sigma, \mathcal{M}_k, p) \leq \epsilon$, and we obtain a contradiction). The theorem follows by applying Lemma 4.5 and Observation 3.3. ◀

5 The upper bound for testing under the uniform distribution

In this section we prove the upper bound of Theorem 1.3, which is restated next.

▶ **Theorem 1.3** *The sample complexity of one-sided error sample-based testing of k-modality under the uniform distribution is $\Theta\left(\sqrt{\frac{kn}{\epsilon}}\right)$. The lower bounds holds for any $\epsilon < 1/4$ and $k \leq \epsilon n$.*

Referring to the notations introduced in Section 2, when p is the uniform distribution over $[n]$, we shall use the shorthand $\text{dist}(\sigma, \mathcal{P})$ for $\text{dist}(\sigma, \mathcal{P}, p)$, $I_n(s)$ for $I_n(s,p)$, and $\text{del}(\sigma, \mathcal{P})$ for $\text{del}(\sigma, \mathcal{P}, p)$. Actually, rather than working with the weighted/normalized deletion distance $\text{del}(\sigma, \mathcal{P})$, it will be convenient to work with the absolute deletion distance $\text{Del}(\sigma, \mathcal{P}) = \text{del}(\sigma, \mathcal{P}) \cdot n$. Namely, $\text{Del}(\sigma, \mathcal{P})$ is the minimum *size* of a subset $D \subseteq [n]$ such that $\sigma_{|[n] \setminus D} \in \mathcal{P}$ (where for hereditary properties, such a subset always exists).

For the sake of the analysis, it will be useful to analyze the sample complexity of testing k-modality (with one-sided error and under the uniform distribution) when the sample is selected according to the Poisson distribution. Recall that the Poisson distribution $\text{Poi}(\lambda)$ takes value $x \in \mathbb{N}$ with probability $e^{-\lambda}\lambda^x/x!$. The next definition is analogous to Definition 2.2.

▶ **Definition 5.1.** *For positive integers n and s, we use $I_n^{\text{Poi}}(s)$ to denote the random variable consisting of a subset of $[n]$ such that for each $i \in [n]$ we have $\Pr[i \in I_n^{\text{Poi}}(s)] = \Pr[\text{Poi}(s/n) \neq 0] = 1 - e^{-s/n}$.*

The following lemma is directly implied by [11, Lemma 2.2] (which in turn refers to [47]).

▶ **Lemma 5.2.** *For any property $\mathcal{P}$, positive integers n and s, and sequence σ of length n,*

$$\Pr[\sigma_{|I_n(s)} \notin \mathcal{P}] \geq \Pr[\sigma_{|I_n^{\text{Poi}}(s/2)} \notin \mathcal{P}] - \frac{4}{s}\,.$$

We next introduce a definition and a simple claim.

▶ **Definition 5.3.** *An ascent/descent (x, y) is said to* start *at x and* end *at y.*
We define the first *ascent (descent) in a sequence σ to be the ascent (descent) that ends first. In case that there are multiple such ascents (descents), choose the one that starts first.*
For an integer $r > 1$, we recursively define the r^{th} ascent (descent) *in σ to be the first ascent (descent) in the subsequence obtained from σ by deleting the first $r-1$ ascents (descents).*

Recall that we use the notations $\Uparrow$-pair for an ascent and $\Downarrow$-pair for a descent.

▷ Claim 5.4. Let σ be a sequence of length n, $t \geq 2$ an integer, and $\Updownarrow \in \{\Uparrow, \Downarrow\}$. Then σ contains at least $\frac{1}{2}\mathrm{Del}(\sigma, \mathcal{F}_t^{\Updownarrow})$ disjoint $\Updownarrow$-pairs.

Proof. We prove that at least $\frac{1}{2}\mathrm{Del}(\sigma, \mathcal{F}_t^{\Updownarrow})$ disjoint $\Updownarrow$-pairs exist. Delete the first $\Updownarrow$-pair (as defined in Definition 5.3), if it exists. Continue to delete the second $\Updownarrow$-pair, third $\Updownarrow$-pair, etc. as long as possible. Assume that after deleting the y^{th} $\Updownarrow$-pair, no $\Updownarrow$-pair remains. Therefore, by deleting $2y$ symbols from σ we turned it into a sequence in $\mathcal{F}_2^{\Uparrow}$, and hence also in $\mathcal{F}_t^{\Uparrow}$ (for any $t \geq 2$). Thus $y \geq \frac{1}{2}\mathrm{Del}(\sigma, \mathcal{F}_t^{\Uparrow})$, so at least $\frac{1}{2}\mathrm{Del}(\sigma, \mathcal{F}_t^{\Uparrow})$ disjoint $\Updownarrow$-pairs exist. ◁

The next structural claim is a corollary of Claim 4.1.

▷ Claim 5.5. Let σ be a sequence of length n, $t \geq 2$ an integer, $\Updownarrow \in \{\Uparrow, \Downarrow\}$, and y an integer smaller than $\frac{1}{2}\mathrm{Del}(\sigma, \mathcal{F}_t^{\Updownarrow})$. Let $m \in [n]$ be the index such that the y^{th} $\Updownarrow$-pair in σ ends at m. Then $\mathrm{Del}(\sigma, \mathcal{F}_t^{\Updownarrow}) \leq 2y + \mathrm{Del}(\sigma_{m+1}, \ldots, \sigma_n, \mathcal{F}_{t-1}^{\mathrm{inv}(\Updownarrow)})$.

We are now ready to state and prove the lemma that is the heart of our upper bound argument. As opposed to the upper-bound argument for the distribution-free case (Lemma 4.5), In the proof of Lemma 5.6 we do not apply the structural claim described above to "break" σ into predetermined subsequences and then consider the task of testing monotonicity for each of them. Instead, the subsequences are determined by a process that traverses the sequence while selecting the sample "on the fly", and gathering evidence against monotonicity of subsequences.

▶ **Lemma 5.6.** *Let $\epsilon > 0$, $t \geq 2$ an integer, and $\Updownarrow \in \{\Uparrow, \Downarrow\}$. If $\mathrm{Del}(\sigma, \mathcal{F}_t^{\Updownarrow}) > \epsilon n$, then for $s = \Theta(\sqrt{\frac{tn}{\epsilon}})$ and $Q = I_n^{\mathrm{Poi}}(s)$ we have that*

$$\Pr\left[\sigma_{|Q} \notin \mathcal{F}_t^{\Updownarrow}\right] \geq \frac{5}{6}\,.$$

Proof. Denote $\Delta = \mathrm{Del}(\sigma, \mathcal{F}_t^{\Updownarrow})$, so that $\Delta > \epsilon n$. Let $s = 20\sqrt{\frac{tn}{\epsilon}}$ and consider the random variable $Q = I_n^{\mathrm{Poi}}(s)$. We shall prove that $\Pr[\sigma_{|Q} \notin \mathcal{F}_t^{\Updownarrow}] \geq 5/6$.

Let $r = t - 1$. For $2 \leq u \leq t$, we define $\hat{\mathcal{F}}_u$ as follows.

$$\hat{\mathcal{F}}_u = \begin{cases} \mathcal{F}_u^{\Updownarrow} & t-u \text{ is even} \\ \mathcal{F}_u^{inv(\Updownarrow)} & t-u \text{ is odd} \end{cases}$$

We define a process that given a sample Q, tries to find evidence that $\sigma_{|Q}$ does not belong to $\mathcal{F}_t^{\Updownarrow}$. Following this, we analyze the probability that Q is such that the process succeeds.

To be precise, our process aims to find indices $a_0 < b_0 < a_1 < b_1 < \cdots < a_r < b_r$ and numbers $y_0, \ldots, y_r$, such that the following holds for every $0 \leq i \leq r$:

- **Property 1:** If i is odd, then (a_i, b_i) is a $\Updownarrow$-pair, and if i is even and $i > 0$, then (a_i, b_i) is an $\mathrm{inv}(\Updownarrow)$-pair.
- **Property 2:** $\mathrm{Del}(\sigma_{b_i+1} \ldots \sigma_n, \hat{\mathcal{F}}_{t-i}) \geq \Delta - 2\sum_{j=0}^{i} y_j$.

Notice that if the process succeeds, then in particular $((a_1, b_1), \ldots, (a_r, b_r))$ is an r-$\Updownarrow$-pair sequence, which implies by Observation 3.5 that $\sigma_{|Q} \notin \mathcal{F}_t^{\Updownarrow}$.

The process initializes $a_0 = -1$, $b_0 = 0$ and $y_0 = 0$, so that for $i = 0$ Property 1 holds trivially and Property 2 holds by Δ's definition.

For each $\ell \in [r]$, we henceforth condition on the process having succeeded in finding indices $a_0 < b_0 < \cdots < a_{\ell-1} < b_{\ell-1}$ and $y_0, \ldots, y_{\ell-1}$ such that Property 1 and Property 2 both hold for every $i \leq \ell - 1$. We now try to find a_ℓ, b_ℓ and y_ℓ such that $b_{\ell-1} < a_\ell < b_\ell$, and both properties hold for $i = \ell$ as well. We refer to this attempt as the ℓ^{th} step of the process. For convenience, we assume that ℓ is odd (the other case is analogous). According to our assumption, since Property 2 holds for $i = \ell - 1$, we have that

$$\mathrm{Del}(\sigma_{b_{\ell-1}+1} \dots \sigma_n, \mathcal{F}^{\Updownarrow}_{t-\ell+1}) \geq \Delta - 2\sum_{j=0}^{\ell-1} y_j \ . \tag{3}$$

By Claim 5.4, there exist at least $\frac{1}{2}\mathrm{Del}(\sigma_{b_{\ell-1}+1} \dots \sigma_n, \mathcal{F}^{\Updownarrow}_{t-\ell+1})$ disjoint $\Updownarrow$-pairs in $\sigma_{b_{\ell-1}+1} \dots \sigma_n$, which by Equation (3) is at least $\frac{1}{2}\Delta - \sum_{j=0}^{\ell-1} y_j$. Consider the first $\Updownarrow$-pair, second $\Updownarrow$-pair, etc. in $\sigma_{b_{\ell-1}+1}, \dots, \sigma_n$, as defined in Definition 5.3. Assume that the first $\Updownarrow$-pair among them that belongs to Q is the q^{th} $\Updownarrow$-pair. Then we set $y_\ell = q$, a_ℓ to be the first index of this pair, and b_ℓ to be the last index of this pair (if no such pair belongs to Q, set $y_j = a_j = b_j = \infty$ for every $\ell \leq j \leq r$). If $y_\ell < \frac{1}{2}\Delta - \sum_{j=0}^{\ell-1} y_j$, then by using Claim 5.5 with $y = y_\ell$ and $m = b_\ell$, we infer that $\mathrm{Del}(\sigma_{b_\ell+1} \dots \sigma_n, \mathcal{F}^{\mathrm{inv}(\Updownarrow)}_{r-\ell}) \geq \Delta - 2\sum_{j=0}^{\ell} y_j$ (notice that the last sum now includes y_ℓ).

We conclude that if

$$y_\ell < \frac{1}{2}\Delta - \sum_{j=0}^{\ell-1} y_j \text{ for every } \ell \in [r] \ , \tag{4}$$

then the process succeeds and thus $\sigma_{|Q} \notin \mathcal{F}^{\Updownarrow}_t$. As $y_\ell \geq 0$ for every $\ell \in [r]$ and as $y_0 = 0$, the condition in Equation (4) is equivalent to

$$\sum_{j=1}^{r} y_j < \frac{1}{2}\Delta \ . \tag{5}$$

We now turn to analyze the probability that the condition in Equation (5) holds. Notice that by the definition of $Q = I^{\mathrm{Poi}}_n(s)$, each one of the pairs that were considered during the ℓ^{th} step of the process was sampled with probability ρ^2 for $\rho = 1 - e^{-\frac{s}{n}}$, independent of the others. Observe that from its construction, y_ℓ (for every $\ell \in [r]$) is distributed very similarly to a geometric random variable with parameter ρ^2. To be precise, for every $\ell \in [r]$ let z_ℓ be an i.i.d. geometric random variable with parameter ρ^2, and define $h_\ell = z_\ell$ if z_ℓ is no bigger than the number of pairs that were considered during the ℓ^{th} step, and $h_\ell = \infty$ otherwise. Then h_ℓ and y_ℓ have the same probability distribution. Therefore,

$$\begin{aligned}\Pr_Q[\sigma_{|Q} \notin \mathcal{F}^{\Updownarrow}_t] \ &\geq \ \Pr_Q\left[\sum_{j=1}^{r} y_j < \frac{1}{2}\Delta\right] \ = \ \Pr_{Q,\{z_\ell\}_{\ell\in[r]}}\left[\sum_{j=1}^{r} h_j \ < \ \frac{1}{2}\Delta\right] \\ &\geq \ \Pr_{\{z_\ell\}_{\ell\in[r]}}\left[\sum_{j=1}^{r} z_j < \frac{1}{2}\Delta\right] \ \geq \ 1 - \Pr_{\{z_\ell\}_{\ell\in[r]}}\left[\sum_{j=1}^{r} z_j > \frac{1}{2}\epsilon n\right] \ .\end{aligned}$$

Since $z_\ell \sim G(\rho^2)$ for every $\ell \in [r]$ and they are mutually independent, we know by Fact A.4 that their sum distributes as a negative binomial random variable, and that $\Pr_{\{z_\ell\}_{\ell\in[r]}}[\sum_{j=1}^{r} z_j > \frac{1}{2}\epsilon n] = \Pr[\mathrm{Bin}(\frac{1}{2}\epsilon n, \rho^2) < r]$. Let $\mu = \mathbb{E}[\mathrm{Bin}(\frac{1}{2}\epsilon n, \rho^2)] = \frac{1}{2}\epsilon n \rho^2$, and recall that $\rho = 1 - e^{-\frac{s}{n}}$. Applying Fact A.1 we get $\rho \geq \frac{s}{2n}$, so $\mu \geq \frac{\epsilon s^2}{8n}$. For $s = 20\sqrt{\frac{tn}{\epsilon}}$, the expected value of μ is at least $50t$, which is at least $10r$. Using a tail bound for the binomial distribution (Fact A.3) we conclude that

$$\Pr\left[\frac{1}{2}\mathrm{Bin}(\epsilon n, \rho) < r\right] \leq \Pr\left[\frac{1}{2}\mathrm{Bin}(\epsilon n, \rho) < \frac{1}{10}\mu\right] \leq e^{-\frac{(\frac{9}{10})^2\mu}{2}} \leq e^{-\frac{2}{5}\mu} \leq e^{-\frac{2}{5}\cdot 50t} \leq e^{-20} < \frac{1}{6} \ ,$$

which means that $\Pr_Q\left[\sigma_{|Q} \notin \mathcal{F}^{\Updownarrow}_t\right] \geq 5/6$, as claimed. ◀

The upper bound in Theorem 1.3 follows by combining Observation 3.7 with Lemma 5.6 and Observation 3.3 (in an analogous fashion to what was shown in the proof of Theorem 1.1).

References

1 Nir Ailon and Bernard Chazelle. Information theory in property testing and monotonicity testing in higher dimensions. *Information and Computation*, 204(11):1704–1717, 2006.

2 Nir Ailon, Bernard Chazelle, Ding Liu, and C. Seshadhri. Estimating the distance to a monotone function. *Random Structures and Algorithms*, 31(3):371–383, 2007.

3 Maria Balcan, Eric Blais, Avrim Blum, and Liu Yang. Active property testing. In *Proceedings of the Fifty-Third Annual Symposium on Foundations of Computer Science (FOCS)*, pages 21–30, 2012.

4 Roksana Baleshzar, Deeparnab Chakrabarty, Ramesh Krishnan, S. Pallavoor, and Sofya Raskhodnikova. Optimal unateness testers for real-valued functions. In *Automata, Languages and Programming: Forty-fourth International Colloquium (ICALP)*, pages 5:1–5:14, 2017.

5 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 the Theory of Computing (STOC)*, pages 381–390, 2004.

6 Tugkan Batu, Ronitt Rubinfeld, and Patrick White. Fast approximate PCPs for multidimensional bin-packing problems. *Information and Computation*, 196(1):42–56, 2005.

7 Aleksandra Belovs and Eric Blais. A polynomial lower bound for testing monotonicity. In *Proceedings of the Fourty-Eighth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 1021–1032, 2016.

8 Omri Ben-Eliezer. Testing local properties of arrays. In *Proceedings of the Tenth Innovations in Theoretical Computer Science conference (ITCS)*, pages 11:1–11:20, 2019.

9 Omri Ben-Eliezer and Clément L. Canonne. Improved bounds for testing forbidden order patterns. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2093–2112, 2018.

10 Omri Ben-Eliezer, Clément L. Canonne, Shoham Letzter, and Erik Waingarten. Finding monotone patterns in sublinear time. In *Proceedings of the Sixteeth Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1469–1494, 2019.

11 Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. Testing convexity of figures under the uniform distribution. In *Proceedings of the Thirty-Second International Symposium on Computational Geometry (SoCG)*, pages 17:1–17:15, 2016.

12 Arnab Bhattacharyya, Elena Grigorescu, Kyomin Jung, Sofya Raskhodnikova, and David P. Woodruff. Transitive-closure spanners. *SIAM Journal on Computing*, 41(6):1380–1425, 2012.

13 Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. A $o(d) \cdot \text{polylog} n$ monotonicity tester for Boolean functions over the hypergrid $[n]^d$. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2133–2151, 2018.

14 Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Domain reduction for monotonicity testing: A $o(d)$ tester for Boolean functions in d-dimensions. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1975–1994, 2020.

15 Eric Blais, Sofya Raskhodnikova, and Grigory Yaroslavtsev. Lower bounds for testing properties of functions over hypergrid domains. In *Proceedings of the Twenty-Ninth IEEE Annual Conference on Computational Complexity (CCC)*, pages 309–320, 2014.

16 Jop Briet, Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Monotonicity testing and shortest-path routing on the cube. *Combinatorica*, 32(1):35–53, 2012.

17 Clément L. Canonne, Elena Grigorescu, Siyao Guo, Akash Kumar, and Karl Wimmer. Testing k-monotonicity. In *Proceedings of the Eighth Innovations in Theoretical Computer Science conference (ITCS)*, pages 29:1–29:21, 2017.

18 Deeparnab Chakrabarty, Kashyap Dixit, Madhav Jha, and C. Seshadhri. Property testing on product distributions: Optimal testers for bounded derivative properties. *ACM Transactions on Algorithms*, 13(2):20:1–20:30, 2017.

19 Deeparnab Chakrabarty and C. Seshadhri. Optimal bounds for monotonicity and Lipschitz testing over hypercubes and hypergrids. In *Proceedings of the Fourty-Fifth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 419–428, 2013.

20 Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory of Computing*, 10:453–464, 2014.

21 Deeparnab Chakrabarty and C. Seshadhri. An $o(n)$ monotonicity tester for Boolean functions over the hypercube. *SIAM Journal on Computing*, 45(2):461–472, 2016.

22 Deeparnab Chakrabarty and C. Seshadhri. Adaptive Boolean monotonicity testing in totla influence time. In *Proceedings of the Tenth Innovations in Theoretical Computer Science conference (ITCS)*, pages 20:1–20:7, 2019.

23 Xi Chen, Anindya De, Rocco A. Servedio, and Li-Yang Tan. Boolean function monotonicity testing requires (almost) $n^{1/2}$ non-adaptive queries. In *Proceedings of the Fourty-Seventh Annual ACM Symposium on the Theory of Computing (STOC)*, pages 519–528, 2015.

24 Xi Chen, Rocco A. Servedio, and Li-Yang Tan. New algorithms and lower bounds for monotonicity testing. In *Proceedings of the Fifty-Fifth Annual Symposium on Foundations of Computer Science (FOCS)*, pages 286–295, 2014.

25 Xi Chen and Erik Waingarten. Testing unateness nearly optimally. In *Proceedings of the Fifteeth-First Annual ACM Symposium on the Theory of Computing (STOC)*, pages 547–558, 2019.

26 Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond Talagrand functions: New lower bounds for testing monotonicity and unateness. In *Proceedings of the Fourty-Ninth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 523–536, 2017.

27 Xi Chen, Erik Waingarten, and Jinyu Xie. Boolean unateness testing with $\tilde{O}(n^{3/4})$ adaptive queries. In *Proceedings of the Fifty-Eighth Annual Symposium on Foundations of Computer Science (FOCS)*, pages 868–879, 2017.

28 Ilias Diakonikolas and Daniel Kane. A new approach for testing properties of discrete distributions. In *Proceedings of the Fifty-Seventh Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694, 2016. `doi:10.1109/FOCS.2016.78`.

29 Yevgeniy Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved bounds for testing monotonicity. In *Proceedings of the Third International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 97–108, 1999.

30 Funda Ergün, Sampath Kannan, Ravi Kumar, Ronitt Rubinfeld, and Mahesh Viswanathan. Spot-checkers. In *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 259–268, January 1998.

31 Shahar Fattal and Dana Ron. Approximating the distance to monotonicity in high dimensions. *ACM Transactions on Algorithms*, 6(3):52:1–52:37, 2010.

32 Eldar Fischer. On the strength of comparisons in property testing. *Information and Computation*, 189(1):107–116, 2004.

33 Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on the Theory of Computing (STOC)*, pages 474–483, 2002.

34 Oded Goldreich. The uniform distribution is complete with respect to testing identity to a fixed distribution. ECCC TR16-015, 2016.

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

36 Oded Goldreich and Dana Ron. On sample-based testers. *ACM Transactions on Computing Theory*, 8(2):7:1–7:54, 2016.

37 Shirley Halevy and Eyal Kushilevitz. Distribution-free property testing. *SIAM Journal on Computing*, 37(4):1107–1138, 2007.

38 Shirley Halevy and Eyal Kushilevitz. Testing monotonicity over graph products. *Random Structures and Algorithms*, 33(1):44–67, 2008.

39 Michael Kearns and Dana Ron. Testing problems with sub-learning sample complexity. *Journal of Computer and System Sciences*, 61(3):428–456, 2000.

40 Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and Boolean isoperimentric-type theorems. *SIAM Journal on Computing*, 47(6):2238–2276, 2018.

41 Subhash Khot and Igor Shinkar. An $\tilde{O}(n)$ queries adaptive tester for unateness. In *Proceedings of the Twentieth International Workshop on Randomization and Computation (RANDOM)*, pages 37:1–37:7, 2016.

42 Ramesh Krishnan, S. Pallavoor, Sofya Raskhodnikova, and Nithin M. Varma. Paramaterized property testing of functions. *ACM Transactions on Computing Theory*, 9(4):17:1–17:19, 2018.

43 Ramesh Krishnan, S. Pallavoor, Sofya Raskhodnikova, and Erik Waingarten. Approximating the distance to monotonicity of Boolean functions. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1995–2009, 2020.

44 Ilan Newman, Yuri Rabinovich, Deepak Rajendraprasad, and Christian Sohler. Testing for forbidden order patterns in an array. *Random Structures and Algorithms*, 55(2):402–426, 2019.

45 Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006.

46 Sofya Raskhodnikova. Monotonicity testing. Master's thesis, MIT, 1999. Master's thesis.

47 Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM Journal on Computing*, 39(3):813–842, 2009.

48 Wojciech Szpankowski. *Average Case Analysis of Algorithms on Sequences*. Wiley-Interscience, New York, 2001.

A Basic facts and claims

▶ **Fact A.1** (Exponential inequality). *For* $0 \leq p \leq 1$*,*

$$\frac{1}{2}p \leq p - \frac{1}{2}p^2 \leq 1 - e^{-p} \leq p \;.$$

▶ **Fact A.2** (Markov's inequality). *If* X *is a nonnegative random variable and* $a > 0$*, then*

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[x]}{a} \;.$$

▶ **Fact A.3** (Tail bound for the binomial distribution). *Let* μ *denote the expected value of* $\mathrm{Bin}(n, p)$*, i.e.,* $\mu = np$*. Then for* $\delta \in [0, 1]$*,*

$$\Pr[\mathrm{Bin}(n, p) \leq (1 - \delta)\mu] \leq e^{-\frac{\delta^2\mu}{2}} \;,$$

▶ **Fact A.4** (Sums of independent geometrically distributed random variables). *Let* $W(n, p)$ *denote the sum of* n *independent geometric random variables with parameter* p*. The variable* $W(n, p)$ *is said to have* negative binomial distribution*, and it satisfies*

$$\Pr[W(n, p) \leq m] = \Pr[\mathrm{Bin}(m, p) \geq k] \;.$$

B Proof of Observation 3.6

Since $\mathcal{M}_k$ is a hereditary property, $\text{del}(\sigma, \mathcal{M}_k, p) \le \text{dist}(\sigma, \mathcal{M}_k, p)$. We next show that $\text{dist}(\sigma, \mathcal{M}_k, p) \le \text{del}(\sigma, \mathcal{M}_k, p)$. By the definition of $\text{del}(\sigma, \mathcal{M}_k, p)$, there exists a subset $D \subseteq [n]$ such that $\sigma_{[n]\setminus D} \in \mathcal{M}_k$ and $p(D) = \text{del}(\sigma, \mathcal{M}_k, p)$. If there is more than one such subset, we select one with minimal size, so that $D \neq [n]$. We next define a sequence τ such that $\tau \in \mathcal{M}_k$, and such that $\tau_i = \sigma_i$ for every $i \in [n] \setminus D$, implying that $\text{dist}(\sigma, \mathcal{M}_k, p) \le \text{dist}(\sigma, \tau, p) = p(D) \le \text{del}(\sigma, \mathcal{M}_k, p)$, as desired. We define τ as follows: For each $i \in [n]$, if $i \notin D$, then $\tau_i = \sigma_i$, and if $i \in D$, then $\tau_i = \sigma_j$ for $j \notin D$ such that $|j - i|$ is minimized (breaking ties arbitrarily). To verify that $\tau \in \mathcal{M}_k$, assume, contrary to the claim, that $\tau \notin \mathcal{M}_k$. By Observation 3.3, this means that for $\Updownarrow \in \{\Uparrow, \Downarrow\}$ and $t = k + 3$, there is a t-$\Updownarrow$ subsequence $(x_1, \ldots, x_t)$ with respect to τ. But by the definition of τ, this implies that there exists a t-$\Updownarrow$ subsequence $(x'_1, \ldots, x'_t)$ with respect to $\sigma_{[n]\setminus D}$, and we have reached a contradiction.

C The lower bounds

In this section we prove Theorem 1.3 and the lower bound of Theorem 1.2. Given n, k and ϵ, we construct a sequence and a corresponding probability distribution that are determined by a parameter m. The two lower bounds differ in the setting of this parameter.

Let ϵ, k satisfy $\epsilon < 1/4$ and $k < n/4 - 1$, and let m be an integer satisfying $2k \le m < n/2$. Consider the sequence $\sigma = 2, 1, 4, 3, \ldots, 2m, 2m-1, 3m, 3m, \ldots, 3m$ (where the value $3m$ appears $n - 2m$ times and was chosen as an arbitrary value greater than $2m$). Define $p : [n] \to [0, 1]$ by $p(i) = \rho = \frac{2\epsilon}{m}$ for $i \le 2m$ and $p(i) = \frac{1-2m\rho}{n-2m} = \frac{1-4\epsilon}{n-2m}$ for $i > 2m$, so that $\sum_{i\in[n]} p(i) = 1$. We shall show that σ is ϵ-far from being k-modal with respect to p, but the probability that a sample of size $s = \frac{1}{5}\frac{\sqrt{km}}{\epsilon}$, selected according to p, contains a subsequence that is not k-modal, is a small constant.

First, note that σ has exactly m descents: $(1, 2), \ldots, (2m-1, 2m)$, that is: $(2i-1, 2i)$ for every $i \in [m]$. Next, observe that any subsequence of σ with k descents is not k-modal (in fact, $\frac{k}{2} + 2$ descents are sufficient). Thus, by deleting at most $m - k$ indices, it is possible to eliminate at most $m - k$ descents, so that the resulting sequence is not k-modal. Since $p(i) = \rho$ for every $i \in [2m]$, we get that $\text{del}(\sigma, \mathcal{F}_k, p) > (m-k)\rho \ge (m/2)\rho = \epsilon$, where the last inequality follows from the condition $m \ge 2k$. By Observation 3.6, $\text{dist}(\sigma, \mathcal{F}_k, p) > \epsilon$, as claimed.

We now turn to show that the probability of sampling a subsequence of σ that is not k-modal is very low using $s = \frac{1}{5}\frac{\sqrt{km}}{\epsilon}$ samples. We do so by bounding the number of descents in the sampled subsequence, and the proof is a variant of a birthday-paradox argument. Let $q_1, \ldots, q_s$ denote our s samples. For every two different indices $\alpha, \beta \in [s]$, we define the event $E_{\alpha,\beta} = \{(q_\alpha, q_\beta) \text{ is a descent}\}$. Since there are exactly m descents and they are all disjoint, there are m options (each of weight ρ) for a first index in a descent, and given such an index there is exactly one option for the second index. As any two samples are independent, this means that $\Pr[E_{\alpha,\beta}] = m\rho \cdot \rho = m\rho^2$ for every $\alpha, \beta \in [s]$ such that $\alpha \neq \beta$. For every $\alpha, \beta \in [s]$ such that $\alpha \neq \beta$, let $\chi_{\alpha,\beta} = \mathbb{1}_{E_{\alpha,\beta}}$ denote the indicator function of the event $E_{\alpha,\beta}$. Then $X = \sum_{\alpha\neq\beta} \chi_{\alpha,\beta}$ is the number of descents in our sample. Using linearity of expectation, we can calculate its expected value:

$$\mathbb{E}[X] = \sum_{\alpha\neq\beta} \mathbb{E}[\chi_{\alpha,\beta}] = \sum_{\alpha\neq\beta} \Pr[E_{\alpha,\beta}] = s(s-1) \cdot m\rho^2 < \frac{km}{25\epsilon^2} \cdot m \cdot \left(\frac{2\epsilon}{m}\right)^2 < \frac{1}{6}k \,.$$

By Markov's inequality (Fact A.2),

$$\Pr\left[X \geq \frac{1}{2}k\right] \leq \Pr\left[X \geq 3\mathbb{E}[x]\right] \leq \frac{1}{3} .$$

Therefore with probability at least $2/3$ the sampled subsequence contains less than $k/2$ descents and thus must be k-modal. Hence the tester rejects with a small constant probability.

We next choose appropriate values of m and infer Theorem 1.3 and the lower bound of Theorem 1.2.

Proof of Theorem 1.2. Set $m = \lfloor (n-1)/2 \rfloor$, and note that m fulfills its requirements as $k \leq n/4 - 1$ by the premise of the theorem. We conclude that $\Omega(\frac{\sqrt{km}}{\epsilon}) = \Omega(\frac{\sqrt{kn}}{\epsilon})$ samples are necessary for distribution-free one-sided error sample-based testing of k-modality. ◀

Proof of Theorem 1.3, lower bound. We shall assume that $2\epsilon n$ is an integer (otherwise, the analysis is similar but more cumbersome). Set $m = 2\epsilon n$, and note that m fulfills its requirements by the premise of the theorem that $k \leq \epsilon n$ and $\epsilon < 1/4$. We get that both $\rho = 2\epsilon/m = 1/n$ and $\frac{1-4\epsilon}{n-2m} = 1/n$, hence p is the uniform distribution over n. We conclude that $\Omega(\frac{\sqrt{km}}{\epsilon}) = \Omega(\sqrt{\frac{kn}{\epsilon}})$ samples are necessary for one-sided error sample-based testing of k-modality under the uniform distribution. ◀