BIROn - Birkbeck Institutional Research Online

Lachish, Oded and Fischer, E. and Vasudev, Y. (2019) Improving and extending the testing of distributions for shape-restricted properties. Algorithmica 81 (9), pp. 3765-3802. ISSN 0178-4617.

Noname manuscript No.
(will be inserted by the editor)

Improving and extending the testing of distributions for shape-restricted properties

E. Fischer · O. Lachish · Y. Vasudev

the date of receipt and acceptance should be inserted later

Abstract Distribution testing deals with what information can be deduced about an unknown distribution over $\{1,\ldots,n\}$, where the algorithm is only allowed to obtain a relatively small number of independent samples from the distribution. In the extended conditional sampling model, the algorithm is also allowed to obtain samples from the restriction of the original distribution on subsets of $\{1,\ldots,n\}$.

In 2015, Canonne, Diakonikolas, Gouleakis and Rubinfeld unified several previous results, and showed that for any property of distributions satisfying a "decomposability" criterion, there exists an algorithm (in the basic model) that can distinguish with high probability distributions satisfying the property from distributions that are far from it in the variation distance.

We present here a more efficient yet simpler algorithm for the basic model, as well as very efficient algorithms for the conditional model, which until now was not investigated under the umbrella of decomposable properties. Additionally, we provide an algorithm for the conditional model that handles a much larger class of properties.

Our core mechanism is an algorithm for efficiently producing an interval-partition of $\{1,\ldots,n\}$ that satisfies a "fine-grain" quality. We show that with such a partition at hand we can avoid the search for the "correct" partition of $\{1,\ldots,n\}$.

A preliminary version with less refined bounds appeared in the Proceedings of the 34th STACS (2017).

Eldar Fischer
Faculty of Computer Science, Israel Institute of Technology (Technion), Haifa, Israel. E-mail: eldar@cs.technion.ac.il

Oded Lachish
Birkbeck, University of London, London, UK. E-mail: oded@dcs.bbk.ac.uk

Yadu Vasudev
The Institute of Mathematical Sciences, Chennai, India. E-mail: yaduvasudev@gmail.com

Keywords testing · distributions · decomposability · conditional samples

1 Introduction

1.1 Historical background

In most computational problems that arise from modeling real-world situations, we are required to analyze large amounts of data to decide if it has a fixed property. The amount of data involved is usually too large for reading it in its entirety, with respect to both time and storage. In such situations, it is natural to ask for algorithms that can sample points from the data and obtain a significant estimate for the property of interest. The area of *property testing* addresses this issue by studying algorithms that look at a small part of the data and then decide if the object that generated the data has the property or is far (according to some metric) from having the property.

There has been a long line of research, especially in statistics, where the underlying object from which we obtain the data is modeled as a probability distribution. Here the algorithm is only allowed to ask for independent samples from the distribution, and has to base its decision on them. If the support of the underlying probability distribution is large, it is not practical to approximate the entire distribution. Thus, it is natural to study this problem in the context of property testing.

The specific sub-area of property testing that is dedicated to the study of distributions is called *distribution testing*. There, the input is a probability distribution (in this paper the domain is the set $\{1, 2, \ldots, n\}$) and the objective is to distinguish whether the distribution has a certain property, such as uniformity or monotonicity, or is far in the ℓ_1 distance from it. See [7] for a survey about the area of distribution testing.

Testing properties of distributions was studied by Batu et al in [5], where they gave a sublinear query algorithm for testing closeness of distributions supported over the set $\{1, 2, \ldots, n\}$. They did so by extending the idea of collision counting, which was implicitly used for uniformity testing in the work of Goldreich and Ron ([14]). Consequently, various properties of probability distributions were studied, like testing identity with a known distribution ([4, 20, 3, 12]), testing independence of a distribution over a product space ([4, 3]), and testing k-wise independence ([1]).

In recent years, distribution testing has been extended beyond the classical model. A new model called the *conditional sampling model* was introduced. It first appeared independently in [10] and [9]. In the conditional sampling model, the algorithm queries the input distribution μ with a set $S \subseteq \{1, 2, \ldots, n\}$, and gets an index sampled according to μ conditioned on the set S. Notice that if $S = \{1, 2, \ldots, n\}$, then this is exactly like in the standard model. The conditional sampling model allows adaptive querying of μ, since we can choose the set S based on the indexes sampled until now. Chakraborty et al ([9]) and Canonne et al ([10]) showed that testing uniformity can be done with a number of queries not depending on n (the latter presenting an optimal test), and investigated the testing of other properties of distributions. In [9], it is also shown that uniformity can be tested with $\mathrm{poly}(\log n)$ conditional samples by a *non-adaptive* algorithm. In this work, we study the distribution testing problem in the standard (unconditional) sampling model, as well as in the conditional model.

A line of work which is central to our paper, is the testing of distributions for *structure*. The objective is to test whether a given distribution has some structural properties like being monotone ([6]), being a k-histogram ([16,12]), or being log-concave ([3]). Canonne et al ([8]) unified these results to show that if a property of distributions has certain structural characteristics, then membership in the property can be tested efficiently using samples from the distribution. More precisely, they introduced the notion of *L-decomposable* distributions as a way to unify various algorithms for testing distributions for structure. Informally, an L-decomposable distribution μ supported over $\{1, 2, \ldots, n\}$ is one that has an interval partition $\mathcal{I}$ of $\{1, 2, \ldots, n\}$ of size bounded by L, such that for every interval I, either the weight of μ on it is small or the distribution over the interval is close to uniform. A property $\mathcal{C}$ of distributions is L-decomposable if every distribution $\mu \in \mathcal{C}$ is L-decomposable (L is allowed to depend on n). This generalizes various properties of distributions like being monotone, unimodal, log-concave etc. In this setting, their result for a set of distributions $\mathcal{C}$ supported over $\{1, 2, \ldots, n\}$ translates to the following: if every distribution μ from $\mathcal{C}$ is L-decomposable, then there is an efficient algorithm for testing whether a given distribution belongs to the property $\mathcal{C}$.

To achieve their results, Canonne et al ([8]) show that if a distribution μ supported over $[n]$ is L-decomposable, then it is $O(L \log n)$-decomposable where the intervals are of the form $[j2^i + 1, (j+1)2^i]$. This presents a natural approach of computing the interval partition in a recursive manner, by bisecting an interval if it has a large probability weight and is not close to uniform. Once they get an interval partition, they learn the "flattening" of the distribution over this partition, and check if this distribution is close to the property $\mathcal{C}$. The term "flattening" refers to the distribution resulting from making μ conditioned on any interval of the partition to be uniform. When applied to a partition corresponding to a decomposition of the distribution, the learned flattening is also close to the original distribution. Because of this, in the case where there is a promise that μ is L-decomposable, the above can be viewed as a learning algorithm, where they obtain an explicit distribution that is close to μ. Without the promise it can be viewed as an agnostic learning algorithm. For further elaboration of this connection see [11].

1.2 Results and techniques

In this paper, we extend the body of knowledge about testing L-decomposable properties. We improve upon the previously known bound on the sample complexity, and give much better bounds when conditional samples are allowed. Additionally, for the conditional model, we provide a test for a broader family of properties, that we call *atlas-characterizable* properties.

Our approach differs from that of [8] in the manner in which we compute the interval partition. We show that any partition where most intervals that are not singletons have small probability weight is sufficient to learn the distribution μ, even when it is not the original L-decomposition of μ. We show that if a distribution μ is L-decomposable, then the "flattening" of μ with respect to such a partition is close to μ. It turns out that such a partition can be obtained in "one shot" without resorting to a recursive search procedure.

We obtain a partition as above using the method of *partition pulling* that we develop here. Informally, a pulled partition is obtained by sampling indexes from μ, and taking the partition induced by the samples in the following way: Every sampled index is a singleton interval, and the rest of the partition is composed of the maximal intervals between the sampled indexes. Apart from the obvious simplicity of this procedure, it also has the advantage of providing a partition with a significantly smaller number of intervals, linear in L for a fixed ϵ, and with no dependency on n unless L itself depends on it such that the flattened distribution obtained from the partitioning is ϵ-close to the original distribution. This makes our algorithm more efficient in query complexity than the one of [8] in the unconditional sampling model, and leads to a dramatically small sample complexity in the (adaptive) conditional model.

Another feature of our method is that it provides a partition with small weight intervals also when the distribution is not L-decomposable. This allows us to use the partition in a different manner later on. We provide a test under the conditional query model for the extended class of atlas characterizable properties that we define below, which generalizes both decomposable properties and symmetric properties.

The main common ground between our approach for L-decomposable properties and that of [8] is the method of testing by implicit learning, as defined formally in [13] (see [18]). In particular, the results also provide a means to learn a distribution close to μ if μ satisfies the tested property. A learning algorithm for atlas characterizable properties is not (and cannot be) provided; only an "atlas" of the input distribution, rather than the distribution itself, is learned.

Our result for unconditional testing (Theorem 1) gives a $\sqrt{nL}/\mathrm{poly}(\epsilon)$ query algorithm in the standard (unconditional) sampling model for testing an L-decomposable property of distributions. Our method of finding a good partition for μ using pulled partitions, that we explained above, avoids the $\log n$ factor present in Theorem 3.3 of [8]. The same method enables us to extend our results to the conditional query model, which we present for both adaptive and non-adaptive algorithms. Table 1 summarizes our results and compares them with known lower bounds[1].

[1] The lower bounds for unconditional and non-adaptive conditional testing of L-decomposable properties with $L = 1$ are exactly the lower bounds for uniformity testing; the lower bound for adaptive conditional testing follows easily from the proved existence of

Table 1 Summary of our results

	Result	Known lower bound
L-decomposable	(testing and learning)	
Unconditional	$\sqrt{nL} \cdot \mathrm{polylog}(1/\epsilon)/\epsilon^4$ $\sqrt{nL}\log(L) \cdot \mathrm{polylog}(1/\epsilon)/\epsilon^3$	$\Omega(\sqrt{n}/\epsilon^2)$ for $L = 1$ [17]
Adaptive conditional	$L \cdot \mathrm{polylog}(1/\epsilon)/\epsilon^3$	$\Omega(L)$ for some fixed ϵ [9]
Non-adaptive conditional	$L \cdot \mathrm{polylog}(n)/\mathrm{poly}(\epsilon)$	$\Omega(\log n)$ for $L = 1$ and some fixed ϵ [2]
k-characterized by atlases	(testing)	
Adaptive conditional	$k \cdot \mathrm{polylog}(n)/\mathrm{poly}(\epsilon)$	$\Omega(\sqrt{\log\log n})$ for $k = 1$, and some fixed ϵ [9]

In all the above L and k refer to $L(\Theta(\epsilon), n)$ and $k(\Theta(\epsilon), n)$ respectively, and we make no attempt to optimize the hidden coefficient of ϵ in the parameters, or the powers of all $\mathrm{polylog}(1/\epsilon)$ expressions. We also make no attempt to optimize the powers in the expressions for the non-adaptive conditional model and the properties characterized by atlases, since these rely on tests whose query complexity is likely not optimal.

Up to the power of $\log(1/\epsilon)$ and the aforementioned coefficient of ϵ, our bounds for the unconditional model subsume the $O(\sqrt{nL}(\log(n))^{3/2}/\epsilon^3)$ bound in [8] (the extra $1/2$ power of $\log(n)$ there follows from their theorem being formulated using an "L-splitting" notion that is related to decomposition by an extra $\log(n)$). We know of no previous upper bounds for testing decomposable distribution properties under the conditional models.

2 Preliminaries

We denote the set $\{1, \dots, n\}$ by $[n]$, and denote an interval $\{a, \dots, b\} \subseteq [n]$ by $[a, b]$.

We study the problem of testing properties of probability distributions supported over $[n]$, when we are given samples from the distribution. For two distributions μ and χ, we say that μ is *ϵ-far* from χ if they are far in the ℓ_1 norm, that is, $d(\mu, \chi) = \sum_{i\in[n]} |\mu(i) - \chi(i)| > \epsilon$. For a property of distributions $\mathcal{C}$, we say that μ is *ϵ-far* from $\mathcal{C}$ if for all $\chi \in \mathcal{C}$, $d(\mu, \chi) > \epsilon$.

In addition to the ℓ_1 norm between distributions, we also use the ℓ_∞ norm, $\|\mu - \chi\|_\infty = \max_{i\in[n]} |\mu(i) - \chi(i)|$, and the following measure for uniformity.

Definition 1 For a distribution μ over a domain I, we define the *bias* of μ to be $\mathrm{bias}(\mu) = (\max_{i\in I} \mu(i)/\min_{i\in I} \mu(i)) - 1$.

The following observation is easy and will be used implicitly throughout.

properties that have no sub-linear complexity adaptive conditional tests; finally, the lower bound for properties k-characterized by atlases with $k = 1$ is just a bound for a symmetric property constructed there. About the last one, we conjecture that there exist properties with much higher lower bounds.

Observation 1 *For any two distributions μ and χ over a domain I of size m, $d(\mu,\chi) \le m\|\mu-\chi\|_\infty$. Also, $\|\mu-\mathcal{U}_I\|_\infty \le \frac{1}{m}\mathrm{bias}(\mu)$, where $\mathcal{U}_I$ denotes the uniform distribution over I.*

Proof Follows from the definitions.

We study both the standard model, where the algorithm is given indexes sampled from the distribution, and the model of conditional samples. The conditional model was first studied in the independent works of Chakraborty et al ([9]) and Canonne et al ([10]). We first give the definition of a conditional oracle for a distribution μ.

Definition 2 (Conditional oracle) A conditional oracle for a distribution μ supported over $[n]$ is a black-box that takes as input a set $A \subseteq [n]$, samples a point $i \in A$ with probability $\mu\restriction_A(i) = \mu(i)/\sum_{j\in A}\mu(j)$, and returns i. If $\mu(j) = 0$ for all $j \in A$, then the oracle chooses $i \in A$ uniformly at random.[2]

Now we define conditional distribution testing algorithms. We will define and analyze both *adaptive* and *non-adaptive* algorithms.

Definition 3 An *adaptive* conditional distribution testing algorithm for a property of distributions $\mathcal{C}$, with parameters $\epsilon, \delta > 0$, and $n \in \mathbb{N}$, with query complexity $q(\epsilon,\delta,n)$, is a randomized algorithm with access to a conditional oracle of a distribution μ with the following structure:

- For each $i \in [q]$, at the i^{th} phase, the algorithm generates a set $A_i \subseteq [n]$, based on $j_1, j_2, \cdots, j_{i-1}$ and its internal coin tosses, and calls the conditional oracle with A_i to receive an element j_i, drawn independently of $j_1, j_2, \cdots, j_{i-1}$.
- Based on the received elements $j_1, j_2, \cdots, j_q$ and its internal coin tosses, the algorithm accepts or rejects the distribution μ.

If $\mu \in \mathcal{C}$, then the algorithm accepts with probability at least $1-\delta$, and if μ is ϵ-far from $\mathcal{C}$, then the algorithm rejects with probability at least $1-\delta$.

Definition 4 A *non-adaptive* conditional distribution testing algorithm for a property of distributions $\mathcal{C}$, with parameters $\epsilon, \delta > 0$, and $n \in \mathbb{N}$, with query complexity $q(\epsilon,\delta,n)$, is a randomized algorithm with access to a conditional oracle of a distribution μ with the following structure:

- The algorithm chooses sets $A_1, \ldots, A_q$ (not necessarily distinct) based on its internal coin tosses, and then queries the conditional oracle to respectively obtain $j_1, \ldots, j_q$.
- Based on the received elements $j_1, \ldots, j_q$ and its internal coin tosses, the algorithm accepts or rejects the distribution μ.

If $\mu \in \mathcal{C}$, then the algorithm accepts with probability at least $1-\delta$, and if μ is ϵ-far from $\mathcal{C}$, then the algorithm rejects with probability at least $1-\delta$.

[2] The behavior of the conditional oracle on sets A with $\mu(A) = 0$ is as per the model of Chakraborty et al [9]. However, upper bounds in this model also hold in the model of Canonne et al [10], and most known lower bounds can be easily converted to it.

2.1 Large deviation bounds

The following large deviation bounds will be used in the analysis of our algorithms throughout the rest of the paper.

Lemma 1 (Chernoff bounds) *Let $X_1, X_2, \ldots, X_m$ be independent random variables taking values in $\{0,1\}$. Let $X = \sum_{i\in[m]} X_i$. Then for any $\delta \in (0,1]$, we have the following bounds.*

1. $\Pr[X > (1+\delta)E[X]] \le \exp\left(\frac{-\delta^2 E[X]}{3}\right)$.
2. $\Pr[X < (1-\delta)E[X]] \le \exp\left(\frac{-\delta^2 E[X]}{2}\right)$.

If $\delta \ge 1$, then $\Pr[X \ge (1+\delta)E[X]] < \exp(\frac{-\delta E[X]}{3})$.

Lemma 2 (Hoeffding bounds [15]) *Let $X_1, X_2, \ldots, X_m$ be independent random variables such that $0 \le X_i \le 1$ for $i \in [m]$, and let $\bar{X} = \frac{1}{m}\sum_{i\in[m]} X_i$. Then $\Pr\left[|\bar{X} - E[\bar{X}]| > \epsilon\right] \le 2\exp\left(-2m\epsilon^2\right)$*

2.2 Basic distribution procedures

The following is a folklore result about learning any distribution supported over $[n]$, that we prove here for completeness.

Lemma 3 (Folklore) *Let μ be a distribution supported over $[n]$. By using $\frac{2(n+\log(2/\delta))}{\epsilon^2}$ unconditional samples from μ, we can obtain an explicit distribution μ' supported on $[n]$ such that, with probability at least $1-\delta$, $d(\mu,\mu') \le \epsilon$.*

Proof Take $m = \frac{2(n+\log(2/\delta))}{\epsilon^2}$ samples from μ, and for each $i \in [n]$, let m_i be the number of times i was sampled. Define $\mu'(i) = m_i/m$. Now, we show that $\max_{S\subseteq[n]} |\mu(S) - \mu'(S)| \le \epsilon/2$ with probability at least $1-\delta$. The lemma follows from this since the ℓ_1 distance is equal to twice this amount.

For any set $S \subseteq [n]$, let $X_1, X_2, \ldots, X_m$ be random variables such that $X_j = 1$ if the j^{th} sample was in S, and otherwise $X_j = 0$. Let $\bar{X} = \frac{1}{m}\sum_{j\in[m]} X_j$. Then, $\bar{X} = \mu'(S)$ and $E[\bar{X}] = \mu(S)$. By Lemma 2, $\Pr[|\bar{X} - E[\bar{X}]| > \epsilon/2] \le 2e^{-m\epsilon^2/2}$. Substituting for m, we get $\Pr[|\mu'(S) - \mu(S)| > \epsilon/2] \le 2e^{-n-\log(2/\delta)}$. Taking a union bound over all sets, with probability at least $1-\delta$, $|\mu'(S) - \mu(S)| \le \epsilon/2$ for every $S \subseteq [n]$. Therefore, $d(\mu,\mu') \le \epsilon$.

We also have the following simple lemma about learning a distribution under the ℓ_∞ distance.

Lemma 4 *Let μ be a distribution supported over $[n]$. Using $\frac{\log(2n/\delta)}{2\epsilon^2}$ unconditional samples from μ, we can obtain an explicit distribution μ' supported on $[n]$ such that, with probability at least $1-\delta$, $\|\mu-\mu'\|_\infty \le \epsilon$.*

Proof Take $m = \frac{\log(2n/\delta)}{2\epsilon^2}$ samples from μ, and for each $i \in [n]$, let m_i be the number of times i was sampled. For each $i \in [n]$, define $\mu'(i) = m_i/m$.

For an index $i \in [n]$, let $X_1, X_2, \ldots, X_m$ be random variables such that $X_j = 1$ if the j^{th} sample is i, and otherwise $X_j = 0$. Let $\bar{X} = \frac{1}{m}\sum_{j\in[m]} X_j$. Then $\bar{X} = \mu'(i)$ and $E[\bar{X}] = \mu(i)$. By Lemma 2, $\Pr[|\bar{X} - E[\bar{X}]| > \epsilon] \leq 2e^{-2m\epsilon^2}$. Substituting for m, we get that $\Pr[|\mu(i) - \mu'(i)| > \epsilon] \leq \delta/n$. By a union bound over $[n]$, with probability at least $1 - \delta$, $|\mu(i) - \mu'(i)| \leq \epsilon$ for every $i \in [n]$.

Finally, a simple lemma about classifying the probabilities of many sets at once.

Lemma 5 *Given subsets $A_1, \ldots A_s$ of $[n]$ and any $\gamma, \delta > 0$, there is a procedure taking $O(\log(s/\delta)/\gamma)$ many unconditional samples, which with probability $1 - \delta$ provides non-negative integers $k_1, \ldots, k_s$ for which the following holds for every $1 \leq t \leq s$.*

- *If $\mu(A_t) < \gamma/2$ then $k_t = 0$.*
- *Otherwise, k_t is such that $2^{k_t-1}\gamma \leq \mu(A_t) < 2^{k_t+1}\gamma$.*

Proof We take $m = 50\log(s/\delta)/\gamma$ many samples, set $k_t = 0$ for every $1 \leq t \leq s$ such that less than $2\gamma m/3$ samples landed in A_t, and for every other t set k_t such that at least $2^{k_t}\gamma m/3$ but less than $2^{k_t+1}\gamma m/3$ samples landed in A_t. Lemma 1 (with $\delta = 1/3$ in the Chernoff bounds) provides that every A_t is correctly classified with probability at least $1 - \delta/s$, and a union bound concludes the proof.

3 Fine partitions and how to pull them

We define the notion of η-fine partitions of a distribution μ supported over $[n]$, which are central to all our algorithms.

Definition 5 (η-fine interval partition) Given a distribution μ over $[n]$, an *η-fine* interval partition of μ is an interval-partition $\mathcal{I} = (I_1, I_2, \ldots, I_r)$ of $[n]$ such that for all $j \in [r]$, $\mu(I_j) \leq \eta$, except in the case that $|I_j| = 1$. The *length* $|\mathcal{I}|$ of an interval partition $\mathcal{I}$ is the number of intervals in it.

The following Algorithm 1 is the mechanism for pulling a fine partition. It takes independent unconditional samples from μ, makes them into singleton intervals in the interval-partition $\mathcal{I}$, and takes the intervals between these samples as the remaining intervals in $\mathcal{I}$.

Lemma 6 *Let μ, η and δ, be the input to Algorithm 1. Then, with probability at least $1-\delta$, the set of intervals $\mathcal{I}$ returned by Algorithm 1 is an η-fine interval partition of μ of length $O\left(\frac{1}{\eta}\log\left(\frac{1}{\eta\delta}\right)\right)$.*

Algorithm 1: Pulling an η-fine partition

Input: Oracle access to a distribution μ supported over $[n]$, parameters $\eta > 0$ (fineness) and $\delta > 0$ (error probability)

1 take $m = \frac{3}{\eta} \log\left(\frac{3}{\eta\delta}\right)$ unconditional samples from μ
2 arrange the indices sampled in increasing order $i_1 < i_2 < \cdots < i_m$ without repetition and set $i_0 = 0$
3 **for** *each* $j \in [m]$ **do**
4 add the singleton interval $\{i_j\}$ to $\mathcal{I}$
5 **if** $i_j > i_{j-1} + 1$ **then** add the interval $[i_{j-1} + 1, i_j - 1]$ to $\mathcal{I}$
6 **if** $i_m < n$ **then** add the interval $[i_m + 1, n]$ to $\mathcal{I}$
7 **return** $\mathcal{I}$

Proof Let $\mathcal{I}$ be the set of intervals returned by Algorithm 1. The guarantee on the length of $\mathcal{I}$ follows from the number of samples taken in Step 1, noting that $|\mathcal{I}| \leq 2m - 1 = O(m)$.

Let $\mathcal{J}$ be a maximal set of pairwise disjoint intervals in $[n]$, where every $I \in \mathcal{J}$ is a minimal interval for which $\mu(I) \geq \eta/3$. Note that every i for which $\mu(i) \geq \eta/3$ necessarily appears as a singleton interval $\{i\} \in \mathcal{J}$. Also clearly $|\mathcal{J}| \leq 3/\eta$.

We shall first show that, with probability at least $1 - \delta$, the samples taken in Step 1 include an index from every interval $I \in \mathcal{J}$. Afterwards, we show that every interval I' such that $\mu(I') \geq \eta$ contains some interval $I \in \mathcal{J}$. By Steps 2 to 6 of the algorithm, these two assertions imply that an η-fine partition is produced.

Let $I \in \mathcal{J}$. The probability that an index from I is not sampled is at most $(1-\eta/3)^{3\log(3/\eta\delta)/\eta} \leq \delta\eta/3$. By a union bound over all $I \in \mathcal{J}$, with probability at least $1 - \delta$ the samples taken in Step 1 include an index from every interval in $\mathcal{J}$.

Now let I' be an interval such that $\mu(I') \geq \eta$, and assume on the contrary that it contains no interval from $\mathcal{J}$. Clearly it may intersect without containing at most two intervals $I_l, I_r \in \mathcal{J}$. Also, $\mu(I' \cap I_l) < \eta/3$ because otherwise we could have replaced I_l with $I' \cap I_l$ in $\mathcal{J}$, and the same holds for $\mu(I' \cap I_r)$. But this means that $\mu(I' \setminus (I_l \cup I_r)) > \eta/3$, and so we could have added $I' \setminus (I_l \cup I_r)$ (or a subinterval thereof) to $\mathcal{J}$, again a contradiction.

The following is a definition of a variation of a fine partition, where we allow some intervals of small total weight to violate the original requirements.

Definition 6 ((η, γ)-fine partitions) Given a distribution μ over $[n]$, an (η, γ)-*fine* interval partition is an interval partition $\mathcal{I} = (I_1, I_2, \ldots, I_r)$ such that $\sum_{I \in \mathcal{H}_\mathcal{I}} \mu(I) \leq \gamma$, where $\mathcal{H}_\mathcal{I}$ is defined as the set of *violating intervals* $\{I \in \mathcal{I} : \mu(I) > \eta, |I| > 1\}$.

In our applications, γ will be larger than η by a factor of L, which will allow us through the following Algorithm 2 to avoid having an additional $\log(L)$ factor in our complexity bounds for the unconditional and the adaptive tests.

Algorithm 2: Pulling an (η, γ)-fine partition

Input: Oracle access to a distribution μ supported over $[n]$, parameters $\eta > 0$ (fineness), $\gamma > 0$ (allowed exception) and $\delta > 0$ (error probability)

1 take $m = \frac{3}{\eta} \log\left(\frac{5}{\gamma\delta}\right)$ unconditional samples from μ

2 perform Step 2 through Step 6 of Algorithm 1

3 **return** $\mathcal{I}$

Lemma 7 *Let μ, η, γ and δ, be the input to Algorithm 2. Then, with probability at least $1-\delta$, the set of intervals $\mathcal{I}$ returned by Algorithm 2 is an (η, γ)-fine interval partition of μ of length $O\left(\frac{1}{\eta}\log\left(\frac{1}{\gamma\delta}\right)\right)$.*

Proof Let $\mathcal{I}$ be the set of intervals returned by Algorithm 2. The guarantee on the length of $\mathcal{I}$ follows from the number of samples taken in Step 1.

As in the proof of Lemma 6, let $\mathcal{J}$ be a maximal set of pairwise disjoint intervals in $[n]$, where every $I \in \mathcal{J}$ is a minimal interval for which $\mu(I) \geq \eta/3$. Here, also define the set $\mathcal{J}'$ to be the set of maximal intervals in $[n] \setminus \bigcup_{I\in\mathcal{J}} I$. Note that $\mathcal{J} \cup \mathcal{J}'$ is an interval partition of $[n]$. Note also that between every two consecutive intervals of $\mathcal{J}'$ lies an interval of $\mathcal{J}$. Finally, since $\mathcal{J}$ is maximal, all intervals in $\mathcal{J}'$ are of weights less than $\eta/3$.

Let $I \in \mathcal{J}$. The probability that an index from I is not sampled is at most $(1-\eta/3)^{3\log(5/\gamma\delta)/\eta} \leq \delta\gamma/5$. By applying the Markov bound over all $I \in \mathcal{J}$ (along with their weights), with probability at least $1-\delta$ the samples taken in Step 1 include an index from every interval in $\mathcal{J}$ except a subset of them of total weight at most $\gamma/5$. Suppose that the above event occurred.

Recalling the definition $\mathcal{H}_{\mathcal{I}} = \{I \in \mathcal{I} : \mu(I) > \eta, |I| > 1\}$, as in the proof of Lemma 6, every $I' \in \mathcal{H}_{\mathcal{I}}$ must fully contain an interval from $\mathcal{J}$ from which no point was sampled. Moreover, I' may not fully contain intervals from $\mathcal{J}$ from which any points were sampled.

Note furthermore that the weight of any such interval $I' \in \mathcal{H}_{\mathcal{I}}$ is not more than 5 times the total weight of the intervals in $\mathcal{J}$ that it fully contains. To see this, recall that the (at most two) intervals from $\mathcal{J}$ that intersect I' without containing it have intersections of weight not more than $\eta/3$. Also, there may be the intervals of $\mathcal{J}'$ intersecting I', each of weight at most $\eta/3$. However, because there is an interval in $\mathcal{J}$ between any two consecutive intervals of $\mathcal{J}'$, the number of intervals from $\mathcal{J}'$ intersecting I' is at most 1 more than the number of intervals of $\mathcal{J}$ fully contained in I'. Thus the number of intersecting intervals from $\mathcal{J} \cup \mathcal{J}'$ is not more than 5 times the number of fully contained intervals from $\mathcal{J}$, and together with their weight bounds we get the bound on the weight the interval I'. By the above this means that $\sum_{I'\in\mathcal{H}_{\mathcal{I}}} \mu(I') \leq \gamma$.

4 Handling decomposable distributions

We now formally define L-decomposable distributions and properties, following [8].

Definition 7 ((γ, L)-decomposable distributions [8]) For an integer L, a distribution μ supported over $[n]$ is (γ, L)-decomposable, if there exists an interval partition $\mathcal{I} = (I_1, I_2, \ldots, I_\ell)$ of $[n]$, where $\ell \leq L$, such that for all $j \in [\ell]$, at least one of the following holds.

1. $\mu(I_j) \leq \frac{\gamma}{L}$.
2. $\max_{i \in I_j} \mu(i) \leq (1+\gamma) \min_{i \in I_j} \mu(i)$.

The second condition in the definition of a (γ, L)-decomposable distribution is identical to saying that $\mathrm{bias}(\mu\restriction_{I_j}) \leq \gamma$. An L-decomposable property is now defined in terms of all its members being decomposable distributions.

Definition 8 (L-decomposable properties, [8]) For a function $L : (0,1] \times \mathbb{N} \to \mathbb{N}$, we say that a property of distributions $\mathcal{C}$ is L-decomposable, if for every $\gamma > 0$, and $\mu \in \mathcal{C}$ supported over $[n]$, μ is $(\gamma, L(\gamma, n))$-decomposable.

Recall that part of our algorithm for learning such distributions is finding (through pulling) what we referred to as a fine partition. Such a partition may still have intervals where the conditional distribution over them is far from uniform. However, we shall show that for L-decomposable distributions, the total weight of such "bad" intervals is not very high.

The next lemma shows that every fine partition of an (γ, L)-decomposable distribution has only a small weight concentrated on "non-uniform" intervals, and thus it will be sufficient to deal with the "uniform" (small bias) intervals.

Lemma 8 *Let μ be a distribution over $[n]$ which is (γ, L)-decomposable. For every γ/L-fine interval partition $\mathcal{I}' = (I'_1, I'_2, \ldots, I'_r)$ of μ, the following holds.*

$$\sum_{j \in [r] : \mathrm{bias}(\mu\restriction_{I'_j}) > \gamma} \mu(I'_j) \leq 2\gamma.$$

Proof Let $\mathcal{I} = (I_1, I_2, \ldots, I_\ell)$ be the L-decomposition of μ, where $\ell \leq L$. Let $\mathcal{I}' = (I'_1, I'_2, \ldots, I'_r)$ be an interval partition of $[n]$ such that for all $j \in [r]$, $\mu(I'_j) \leq \gamma/L$ or $|I'_j| = 1$.

Any interval I'_j for which $\mathrm{bias}(\mu\restriction_{I'_j}) > \gamma$, is either completely inside an interval I_k such that $\mu(I_k) \leq \gamma/L$, or intersects more than one interval (and in particular $|I'_j| > 1$). There are at most $L-1$ intervals in $\mathcal{I}'$ that intersect more than one interval in $\mathcal{I}$. The sum of the weights of all such intervals is at most γ.

For any interval I_k of $\mathcal{I}$ such that $\mu(I_k) \leq \gamma/L$, the sum of the weights of intervals from $\mathcal{I}'$ that lie completely inside I_k is at most γ/L. Thus, the total weight of all such intervals from $\mathcal{I}'$ is bounded by γ. Therefore, the sum of the weights of intervals I'_j such that $\mathrm{bias}(\mu\restriction_{I'_j}) > \gamma$ is at most 2γ.

To get better bounds for our tests, we will use the counterpart of this lemma for the more general (two-parameter) notion of a fine partition.

Lemma 9 *Let μ be a distribution over $[n]$ which is (γ, L)-decomposable. For every $(\gamma/L, \gamma)$-fine interval partition $\mathcal{I}' = (I'_1, I'_2, \ldots, I'_r)$ of μ, the following holds.*

$$\sum_{j \in [r]:\mathrm{bias}(\mu\restriction_{I'_j})>\gamma} \mu(I'_j) \leq 3\gamma.$$

Proof Let $\mathcal{I} = (I_1, I_2, \ldots, I_\ell)$ be the L-decomposition of μ, where $\ell \leq L$. Let $\mathcal{I}' = (I'_1, I'_2, \ldots, I'_r)$ be an interval partition of $[n]$ such that for a set $\mathcal{H}_{\mathcal{I}'}$ of total weight at most γ, for all $I'_j \in \mathcal{I}' \setminus \mathcal{H}_{\mathcal{I}'}$, $\mu(I'_j) \leq \gamma/L$ or $|I'_j| = 1$.

Exactly as in the proof of Lemma 8, the total weight of intervals $I'_j \in \mathcal{I}' \setminus \mathcal{H}_{\mathcal{I}'}$ for which $\mathrm{bias}(\mu \restriction_{I'_j}) > \gamma$ is at most 2γ. In the worst case, all intervals $I'_j \in \mathcal{H}_{\mathcal{I}'}$ are also such that $\mathrm{bias}(\mu \restriction_{I'_j}) > \gamma$, adding at most γ to the total weight of such intervals.

As previously mentioned, we are not directly learning the actual distribution, but a "flattening" thereof. We next formally define the flattening of a distribution μ with respect to an interval partition $\mathcal{I}$. Afterwards we shall describe its advantages and how it can be learned.

Definition 9 Given a distribution μ supported over $[n]$ and a partition $\mathcal{I} = (I_1, I_2, \ldots, I_\ell)$ of $[n]$, the *flattening* of μ with respect to $\mathcal{I}$ is a distribution $\mu_{\mathcal{I}}$, supported over $[n]$, such that for $i \in I_j$, $\mu_{\mathcal{I}}(i) = \mu(I_j)/|I_j|$.

The following lemma shows that the flattening of any distribution μ, with respect to any interval partition that has only small weight on intervals far from uniform, is close to μ.

Lemma 10 *Let μ be a distribution supported on $[n]$, and let $\mathcal{I} = (I_1, I_2, \ldots, I_r)$ be an interval partition of μ such that $\sum_{j \in [r]:d(\mu\restriction_{I_j}, \mathcal{U}_{I_j})>\gamma} \mu(I_j) \leq \eta$. Then $d(\mu, \mu_{\mathcal{I}}) \leq \gamma + 2\eta$.*

Proof We split the sum $d(\mu, \mu_{\mathcal{I}})$ into parts, one for I_j such that $d(\mu \restriction_{I_j}, \mathcal{U}_{I_j}) \leq \gamma$, and one for the remaining intervals.

For $I_j \in \mathcal{I}$ such that $d(\mu \restriction_{I_j}, \mathcal{U}_{I_j}) \leq \gamma$, we have

$$\sum_{i \in I_j} \left| \mu(i) - \frac{\mu(I_j)}{|I_j|} \right| = \sum_{i \in I_j} \mu(I_j) \left| \mu \restriction_{I_j} (i) - \frac{1}{|I_j|} \right| = \mu(I_j) d(\mu \restriction_{I_j}, \mathcal{U}_{I_j}) \leq \gamma \mu(I_j). \tag{1}$$

For $I_j \in \mathcal{I}$ such that $d(\mu \restriction_{I_j}, \mathcal{U}_{I_j}) > \gamma$, we have

$$\sum_{i \in I_j} \left| \mu(i) - \frac{\mu(I_j)}{|I_j|} \right| = \sum_{i \in I_j} \mu(I_j) \left| \mu \restriction_{I_j} (i) - \frac{1}{|I_j|} \right| \leq 2\mu(I_j) \tag{2}$$

We know that the sum of $\mu(I_j)$ over all I_j such that $d(\mu \restriction_{I_j}, \mathcal{U}_{I_j}) \geq \gamma$ is at most η. Using Equations 2 and 1, and summing up over all the sets $I_j \in \mathcal{I}$, the lemma follows.

The good thing about a flattening (for an interval partition of small length) is that it can be efficiently learned. For this we first make a technical definition and note some trivial observations:

Definition 10 (coarsening) Given μ and $\mathcal{I}$, where $|\mathcal{I}| = \ell$, we define the *coarsening* of μ according to $\mathcal{I}$ to be the distribution $\hat{\mu}_{\mathcal{I}}$ over $[\ell]$ so that $\hat{\mu}_{\mathcal{I}}(j) = \mu(I_j)$ for all $j \in [\ell]$.

Observation 2 *Given a distribution $\hat{\mu}_{\mathcal{I}}$ over $[\ell]$, define $\mu_{\mathcal{I}}$ over $[n]$ by $\mu(i) = \hat{\mu}_{\mathcal{I}}(j_i)/|I_{j_i}|$, where j_i is the index satisfying $i \in I_{j_i}$. This is a distribution, and for any two distributions $\hat{\mu}_{\mathcal{I}}$ and $\hat{\chi}_{\mathcal{I}}$ we have $d(\mu_{\mathcal{I}}, \chi_{\mathcal{I}}) = d(\hat{\mu}_{\mathcal{I}}, \hat{\chi}_{\mathcal{I}})$. Moreover, if $\hat{\mu}_{\mathcal{I}}$ is a coarsening of a distribution μ over $[n]$, then $\mu_{\mathcal{I}}$ is the respective flattening of μ.*

Proof All of this follows immediately from the definitions.

The following lemma shows how learning can be achieved. We will ultimately use this in conjunction with Lemma 10 as a means to learn a whole distribution through its flattening.

Lemma 11 *Given a distribution μ supported over $[n]$ and an interval partition $\mathcal{I} = (I_1, I_2, \ldots, I_\ell)$, using $\frac{2(\ell+\log(2/\delta))}{\epsilon^2}$ unconditional samples from μ, we can obtain an explicit distribution $\mu'_{\mathcal{I}}$, supported over $[n]$, such that $d(\mu_{\mathcal{I}}, \mu'_{\mathcal{I}}) \le \epsilon$ with probability at least $1 - \delta$.*

Proof First, note that an unconditional sample from $\hat{\mu}_{\mathcal{I}}$ can be simulated using one unconditional sample from μ. To obtain it, take the index i sampled from μ, and set j to be the index for which $i \in I_j$. Using Lemma 3, we can now obtain a distribution $\hat{\mu}'_{\mathcal{I}}$, supported over $[\ell]$, such that $d(\hat{\mu}_{\mathcal{I}}, \hat{\mu}'_{\mathcal{I}}) \le \epsilon$ with probability at least $1 - \delta$. To finish, we construct and output $\mu'_{\mathcal{I}}$ as per Observation 2.

5 Weakly tolerant interval uniformity tests

To unify as much as possible our treatment of learning and testing with respect to L-decomposable properties to all three models (unconditional, adaptive-conditional and non-adaptive-conditional), we first define what it means to test a distribution μ for uniformity over an interval $I \subseteq [n]$. The following definition is technical in nature, but it is what we need to use as a building block for our learning and testing algorithms.

Definition 11 (weakly tolerant interval tester) A *weakly tolerant interval tester* is an algorithm $\mathbb{T}$ that takes as input a distribution μ over $[n]$, an interval $I \subseteq [n]$, a maximum size parameter m, a minimum weight parameter γ, an approximation parameter ϵ and an error parameter δ, and satisfies the following.

1. If $|I| \le m$, $\mu(I) \ge \gamma$, and $\mathrm{bias}(\mu\restriction_I) \le \epsilon/100$, then $\mathbb{T}$ accepts with probability at least $1 - \delta$.

2. If $|I| \leq m$, $\mu(I) \geq \gamma$, and $d(\mu\restriction_I, \mathcal{U}_I) > \epsilon$, then $\mathbb{T}$ rejects with probability at least $1 - \delta$.

In all other cases, $\mathbb{T}$ may accept or reject with arbitrary probability.

For our purposes we will use three weakly tolerant interval testers, one for each model.

First, a tester for uniformity which uses unconditional samples, a version of which has already appeared implicitly in [14]. We state below the tester with the best dependence on n and ϵ. We first state it in its original form, where I is the whole of $[n]$, implying that $m = n$ and $\gamma = 1$, and $\delta = 1/3$.

Lemma 12 ([17]) *For the input* $(\mu, [n], n, 1, \epsilon, 1/3)$*, there is a weakly tolerant interval tester using* $O(\sqrt{n}/\epsilon^2)$ *unconditional samples from* μ*.*

The needed adaptation to our purpose is straightforward.

Lemma 13 *For the input* $(\mu, I, m, \gamma, \epsilon, \delta)$*, there is a weakly tolerant interval tester which uses* $O(\sqrt{m}\log(1/\delta)/\gamma\epsilon^2)$ *unconditional samples from* μ*.*

Proof To adapt the tester of Lemma 12 to the general m and γ, we just take samples according to μ and keep from them those samples that lie in I. This simulates samples from $\mu\restriction_I$, over which we employ the original tester. This gives a tester using $O(\sqrt{m}/\gamma\epsilon^2)$ unconditional samples and providing an error parameter of, say, $\delta = 2/5$ (the extra error is due to the probability of not getting enough samples from I even when $\mu(I) \geq \gamma$). To move to a general δ, we repeat this process $O(\log(1/\delta))$ times and take the majority vote.

Next, a tester that uses adaptive conditional samples. For this we use the following tester from [10] (see also [9]). Its original statement does not have the weak tolerance (acceptance for small bias) guarantee, but it is easy to see that the proof there works for the stronger assertion. This time we skip the question of how to adapt the original algorithm from $I = [n]$ and $\delta = 2/3$ to the general parameters here. This is since γ does not matter (due to using adaptive conditional samples), the query complexity is independent of the domain size to begin with, and the move to a general $\delta > 0$ is by standard amplification.

Lemma 14 ([10], see also [9]) *For the input* $(\mu, I, m, \gamma, \epsilon, \delta)$*, there is a weakly tolerant interval tester that adaptively takes* $\mathrm{polylog}(1/\epsilon)\log(1/\delta)/\epsilon^2$ *conditional samples from* μ*.*

Finally, a tester that uses non-adaptive conditional samples. For this to work in our test, it is also very important that the queries do not depend on I (but only on n and γ). We just state here the lemma, the algorithm itself is presented and analyzed in Section 9.

Lemma 15 *For the input* $(\mu, I, m, \gamma, \epsilon, \delta)$*, there is a weakly tolerant interval tester that non-adaptively takes* $\mathrm{poly}(\log n, 1/\epsilon)\log(1/\delta)/\gamma$ *conditional samples from* μ*, in a manner that is additionally independent of the interval* I*.*

5.1 Testing uniformity in parallel

To get a better dependency on ϵ for the unconditional setting, we will use a version of Lemma 13 that "scales" better by allowing us to test several disjoint intervals at once. We formulate it for testing sets rather than just intervals, and not necessarily disjoint ones.

The way we formulate it (obtaining correct answers for all sets rather than just most of them) adds an additional $\log(1/\epsilon)$ in the bound proved in Section 6, but we do not optimize the power of $\log(1/\epsilon)$ in our complexity estimates anyway.

Lemma 16 *For the input $(\mu, A_1, \ldots, A_s, m, \gamma, \epsilon, \delta)$, where $A_1, \ldots, A_s$ are subsets of $[n]$, there is a tester which uses $O(\sqrt{m}\log(s)\log(1/\delta)/\gamma\epsilon^2)$ unconditional samples from μ, and with probability at least $1-\delta$, supplies for all sets answers which satisfy the following.*

1. *If $|A_t| \leq m$, $\mu(A_t) \geq \gamma$, and $\mathrm{bias}(\mu \restriction_{A_t}) \leq \epsilon/100$, then A_t is marked as accepted.*
2. *If $|A_t| \leq m$, $\mu(A_t) \geq \gamma$, and $d(\mu \restriction_{A_t}, \mathcal{U}_{A_t}) > \epsilon$, then A_t is marked as rejected.*
3. *Any other A_t is marked arbitrarily.*

Proof We adapt the tester of Lemma 12 by taking $\Theta(\sqrt{m}\log(s)\log(1/\delta)/\gamma\epsilon^2)$ independent samples from μ, then for each $1 \leq t \leq s$ collecting those samples that lend in A_t, and feeding them to the uniformity test for that set. By lemma 1, with probability at least $\delta/2$ we received at least $\Theta(\sqrt{m}\log(s)\log(1/\delta)/\epsilon^2)$ samples in every A_t whose weight is at least γ, which enables us to test its uniformity with success probability at least $\delta/2s$ if it also satisfies $|A_t| \leq m$. A union bound gives us a success probability of at least $1-\delta$ for all relevant sets at once.

Remark 1 We know of no parallel variant of Lemma 14. The proof of Lemma 15 does lend itself to parallelization, but that endeavor would be futile for the purpose here, because in any case the powers of $\log(n)$ and ϵ in its (current) query complexity are suboptimal.

6 Assessing an interval partition

Through either Lemma 6 or Lemma 7 we know how to construct a fine partition, and then through either Lemma 8 or Lemma 9 respectively we know that if μ is decomposable, then most of the weight is concentrated on intervals with a small bias. However, eventually we would like a test that works for decomposable and non-decomposable distributions alike. For this we need a way to assess an interval partition as to whether it is indeed suitable for learning a distribution. This is done through a weighted sampling of intervals, for which we employ a weakly tolerant tester. The following is the formal description, given as Algorithm 3.

Algorithm 3: Assessing a partition

Input: Oracle access to a distribution μ supported over $[n]$, parameters c, r, an interval partition $\mathcal{I}$ satisfying $|\mathcal{I}| \leq r$, parameters $\epsilon, \delta > 0$, a weakly tolerant interval uniformity tester $\mathbb{T}$ taking input values $(\mu, I, m, \gamma, \epsilon, \delta)$

1 **for** $s = 20\log(1/\delta)/\epsilon$ *times* **do**
2   take an unconditional sample from μ and let $I \in \mathcal{I}$ be the interval that contains it
3   use the tester $\mathbb{T}$ with input values $(\mu, I, n/c, \epsilon/r, \epsilon, \delta/2s)$
4   **if** *test rejects* **then** add I to $\mathcal{B}$
5 **if** $|\mathcal{B}| > 4\epsilon s$ **then** `reject` **else** `accept`

To analyze it, first, for a fine interval partition, we bound the total weight of intervals where the weakly tolerant tester is not guaranteed a small error probability; recall that $\mathbb{T}$ as used in Step 3 guarantees a correct output only for an interval I satisfying $\mu(I) \geq \epsilon/r$ and $|I| \leq n/c$.

Observation 3 *Define* $\mathcal{N}_\mathcal{I} = \{I \in \mathcal{I} : |I| > n/c$ or $\mu(I) < \epsilon/r\}$. *If* $\mathcal{I}$ *is* (η, γ)*-fine, where* $c\eta + \gamma \leq \epsilon$*, then* $\mu(\bigcup_{I \in \mathcal{N}_\mathcal{I}} I) \leq 2\epsilon$.

Proof Intervals in $\mathcal{N}_\mathcal{I}$ must fall into at least one of the following categories.

- Intervals in $\mathcal{H}_\mathcal{I}$, whose total weight is bounded by γ by the definition of a fine partition.
- Intervals whose weight is less than ϵ/r. Since there are at most r such intervals (since $|\mathcal{I}| \leq r$), their total weight is bounded by ϵ.
- Intervals whose size is more than n/c and are not in $\mathcal{H}_\mathcal{I}$. Every such interval is of weight bounded by η (by the definition of a fine partition) and clearly there are no more than c of those, giving a total weight bound of $c\eta$.

Summing these up concludes the proof.

The following "completeness" lemma states that the case of a fine partition of a decomposable distribution, i.e. the case where most intervals exhibit a small bias, is correctly detected.

Lemma 17 *Suppose that* $\mathcal{I}$ *is* (η, γ)*-fine, where* $c\eta + \gamma \leq \epsilon$*. Define* $\mathcal{G}_\mathcal{I} = \{I \in \mathcal{I} : \mathrm{bias}(\mu\restriction_I) \leq \epsilon/100\}$. *If* $\mu(\bigcup_{I \in \mathcal{G}_\mathcal{I}}) \geq 1 - \epsilon$*, then Algorithm 3 accepts with probability at least* $1 - \delta$.

Proof Note by Observation 3 that the total weight of $\mathcal{G}_\mathcal{I} \setminus \mathcal{N}_\mathcal{I}$ is at least $1 - 3\epsilon$. By the Chernoff bound of Lemma 1, with probability at least $1 - \delta/2$ at most $4\epsilon s$ of the intervals drawn in Step 2 fall into $\mathcal{I} \setminus (\mathcal{G}_\mathcal{I} \setminus \mathcal{N}_\mathcal{I})$.

Finally, note that if I as drawn in Step 2 belongs to $\mathcal{G}_\mathcal{I} \setminus \mathcal{N}_\mathcal{I}$, then with with probability at least $1 - \delta/2s$ the invocation of $\mathbb{T}$ in Step 3 will accept it, so by a union bound with probability at least $1 - \delta/2$ all sampled intervals from this set will be accepted. All events occur together and make the algorithm accept with probability at least $1 - \delta$, concluding the proof.

The following "soundness" lemma states that if too much weight is concentrated on intervals where μ is far from uniform in the ℓ_1 distance, then the algorithm rejects. Later we will show that this is the only situation where μ cannot be easily learned through its flattening according to $\mathcal{I}$.

Lemma 18 *Suppose that $\mathcal{I}$ is (η, γ)-fine, where $c\eta + \gamma \leq \epsilon$. Define $\mathcal{F}_{\mathcal{I}} = \{i : \mathcal{I} : d(\mu \restriction_I, \mathcal{U}_I) > \epsilon\}$. If $\mu(\bigcup_{I \in \mathcal{F}_{\mathcal{I}}}) \geq 7\epsilon$, then Algorithm 3 rejects with probability at least $1 - \delta$.*

Proof Note by Observation 3 that the total weight of $\mathcal{F}_{\mathcal{I}} \setminus \mathcal{N}_{\mathcal{I}}$ is at least 5ϵ. By the Chernoff bound of Lemma 1, with probability at least $1 - \delta/2$ at least $4\epsilon s$ of the intervals drawn in Step 2 fall into this set.

Finally, note that if I as drawn in Step 2 belongs to this set, then with with probability at least $1 - \delta/2s$ the invocation of $\mathbb{T}$ in Step 3 will reject it, so by a union bound with probability at least $1 - \delta/2$ all sampled intervals from this set will be rejected. All events occur together and make the algorithm reject with probability at least $1 - \delta$, concluding the proof.

Finally, we present the query complexity of the algorithm. It is presented as polynomial in $\log(1/\delta)$, but this can be made linear easily by first using the algorithm with $\delta = 1/3$, and then repeating it $O(\log(1/\delta))$ times and taking the majority vote. When we use this lemma later on, both r and c will be linear in the decomposability parameter L for a fixed ϵ, and δ will be a fixed constant (we also show later on how to further improve the query complexity for the unconditional model).

Lemma 19 *Algorithm 3 requires $O(q \log(1/\delta)/\epsilon)$ many samples, where $q = q(n/c, \epsilon/r, \epsilon, \delta/2s)$ is the number of samples that the invocation of $\mathbb{T}$ in Step 3 requires.*

In particular, this provides an unconditional sampling algorithm taking $r\sqrt{n/c} \cdot \mathrm{polylog}(1/\epsilon, 1/\delta)/\epsilon^4$ many samples, an adaptive conditional sampling algorithm taking $\mathrm{polylog}(1/\epsilon, 1/\delta)/\epsilon^3$ many samples, or a non-adaptive conditional sampling algorithm taking $r \cdot \mathrm{polylog}(n, 1/\delta)/\mathrm{poly}(\epsilon)$ many samples.

Proof A single (unconditional) sample is taken each time Step 2 is reached, and all other samples are taken by the invocation of $\mathbb{T}$ in Step 3. This makes the total number of samples to be $s(q + 1) = O(q \log(1/\delta)/\epsilon)$.

The bound for each individual sampling model follows by plugging in Lemma 13, Lemma 14 and Lemma 15 respectively. For the last one it is important that the tester makes its queries completely independently of I, as otherwise the algorithm would not have been non-adaptive.

The above is suboptimal for the unconditional model. To improve this algorithm, to be used later in the learning and testing algorithm achieving the first bound for the unconditional model, we just use the parallelization technique of Lemma 16 when we need to test intervals for uniformity.

Lemma 20 *In the setting of unconditional sampling, Algorithm 3 can also be implemented in a way that requires $r\sqrt{n/c} \cdot \mathrm{polylog}(1/\epsilon, 1/\delta)/\epsilon^3$ many samples.*

Proof Instead of performing the main loop of the algorithm using Lemma 13, we first choose s intervals as in Step 2, and then use Lemma 16 to classify all of them at once (if the same interval was chosen more than once, we just "weigh" the answer obtained for it accordingly).

7 Departitioning and trade-off in the unconditional model

In this section we construct and prove an alternative for the assessment procedure under the unconditional sampling model, that will be used for the algorithm achieving the second bound for learning and testing a distribution in this model. The idea is to check the possibility of grouping together neighboring intervals in the fine partition, while assuring the small bias of the distribution over the grouped intervals, in a way that allows us to weaken the minimum probability requirement of the sets that we test for uniformity. This trades a power of ϵ (resulting from the minimum probability requirement) with a factor of $\log(L)$ (as the procedure requires a union bound for testing $O(L^2)$ candidate intervals at once).

We will now define what it means for some $I \subseteq [n]$ to be too "thin" and hence unsuited for testing here.

Definition 12 A subset $I \subseteq [n]$ is called γ-*thin* (with respect to μ), if $\mu(I) < \gamma|I|/n$.

Observation 4 *Given a partition $\mathcal{I} = (I_1, \ldots, I_r)$ of $[n]$, the total probability of all γ-thin intervals in it is bounded by γ.*

Proof This is immediate from $I_1, \ldots, I_r$ being disjoint.

Lemma 16 can be extended to cover sets of varying sizes, as long as they are not thin. Here we also make the "arbitrariness" for intervals that are too light or thin 1-sided towards the negative.

Lemma 21 *For the input $(\mu, A_1, \ldots, A_s, \eta, \gamma, \epsilon, \delta)$, where $A_1, \ldots, A_s$ are subsets of $[n]$ and $\eta \geq \gamma/n$, there is a tester which uses $O(\sqrt{n/\eta\gamma}\log(s/\delta)/\epsilon^2)$ unconditional samples from μ, and with probability at least $1-\delta$, supplies for all intervals answers which satisfy the following.*

1. *If $\mu(A_t) \geq \eta$, A_t is not γ-thin, and* $\mathrm{bias}(\mu\restriction_{A_t}) \leq \epsilon/100$, *then A_t is marked as accepted.*
2. *If $d(\mu\restriction_{A_t}, \mathcal{U}_{A_t}) > \epsilon$, then A_t is marked as rejected.*
3. *Any other A_t is marked arbitrarily.*

Proof We adapt the tester of Lemma 12 by first taking $\Theta(\sqrt{n/\eta\gamma}\log(s/\delta)/\epsilon^2)$ independent samples from μ, and then for each $1 \leq t \leq s$ collecting those samples that lend in A_t By lemma 1, with probability at least $\delta/2$ we received at least $\Theta(\mu(A_t)\cdot\sqrt{n/\eta\gamma}\log(s/\delta)/\epsilon^2)$ samples in every A_t whose weight is at least η (we use here that $\eta \geq \gamma/n$ and hence $\eta\cdot\sqrt{n/\eta\gamma} \geq 1$).

For an A_t which is not γ-thin in addition to having weight at least η, we note that this quantity is at least $\Theta(\sqrt{\eta} \cdot \sqrt{\gamma |A_t|/n} \cdot \sqrt{n/\eta\gamma} \log(s/\delta)/\epsilon^2) = \Theta(\sqrt{|A_t|} \log(s/\delta)/\epsilon^2)$, which enables us to feed these samples to a uniformity test for A_t which has success probability at least $\delta/2s$. A union bound gives us a success probability of at least $1-\delta$ for all relevant sets at once.

To make sure that the second item in the assertion of the lemma is satisfied also for thin sets and sets of low probability, we set the rule that if the number of samples inside A_t is insufficient for a uniformity test with success probability at least $\delta/2s$ (this can be calculated using the known quantity $|A_t|$), then we mark A_t as rejected without invoking the test.

Given an interval partition $\mathcal{I} = (I_1, \ldots, I_r)$ and a set $A \subseteq [r]$, we denote by I_A the union $\bigcup_{j \in A} I_j$. The following lemma is proved similarly to the proof of Lemma 9.

Lemma 22 *Let μ be a distribution over $[n]$ which is (γ, L)-decomposable. For every $(\gamma/L, \gamma)$-fine interval partition $\mathcal{I} = (I_1, I_2, \ldots, I_r)$ of μ, there are disjoint intervals $K_1, \ldots, K_s \subseteq [r]$ where $s \leq L$, such that* $\mathrm{bias}(\mu \restriction_{I_{K_t}}) \leq \gamma$*, $\mu(I_{K_t}) \geq \gamma/L$, I_{K_t} is not γ-thin for every $1 \leq t \leq k$, and $\mu(I_{[r] \setminus \bigcup_{t=1}^{s} K_t}) \leq 5\gamma$.*

Proof Let $\mathcal{I} = (I_1, \ldots, I_\ell)$ denote the (L, γ)-decomposition of μ. For every $J \in \mathcal{I}$ for which $\mathrm{bias}(\mu \restriction_J) \leq \gamma$ we set $K_J = \{j : I_j \subseteq J\}$. We then keep and renumber the sets K_J for which $\mu(I_{K_J}) \geq \gamma/L$, and I_{K_J} is not γ-thin.

As in the proof of Lemma 9, a count shows that the total weight of the intervals above, before discarding the γ-thin intervals and those with $\mu(I_{K_J}) < \gamma/L$, is at least $1-3\gamma$. Since $|\mathcal{I}| \leq L$, discarding the small weight intervals loses not more than an additional γ in the total weight of the remaining intervals, and by Observation 4, discarding the thin intervals loses not more than another additional γ as well.

The above lemma naturally leads to the following definition.

Definition 13 A *grouping sequence for $\mathcal{I}$* is a sequence $\mathcal{K} = (K_0, \ldots, K_s)$ of disjoint subsets partitioning $[r]$, where $K_1, \ldots, K_s$ (but not necessarily K_0) are intervals. We also define $\mathcal{I}_{\mathcal{K}}$ to be the partition $I_{K_0}, \ldots, I_{K_s}$ (note that I_{K_0} is not necessarily an interval in $[n]$, while all other members are intervals).

A grouping sequence $\mathcal{K}$ is called *γ-calm for $\mathcal{I}$*, if $K_1, \ldots, K_s$ satisfy $d(\mu \restriction_{I_{K_t}}, \mathcal{U}_{I_{K_t}}) \leq \gamma$, and $\mu(I_{K_0}) \leq \gamma$.

In turn, this leads to the following observation which will make all of this useful for learning and testing decomposable distributions.

Observation 5 *If $\mathcal{K}$ is γ-calm for $\mathcal{I}$, then μ is 3γ-close to its flattening: $d(\mu, \mu_{\mathcal{I}_{\mathcal{K}}}) \leq 3\gamma$.*

Proof This follows since for every $t > 0$, by definition $\sum_{i \in I_{K_t}} |\mu(i) - \mu_{\mathcal{I}}(i)| \leq \gamma\mu(I_{K_t})$, while $\sum_{i \in I_{K_0}} |\mu(i) - \mu_{\mathcal{I}}(i)| \leq 2\mu(I_{K_0}) \leq 2\gamma$.

Algorithm 4: Departitioning in the unconditional model

Input: Oracle access to a distribution μ supported over $[n]$, parameters L, r, an interval partition $\mathcal{I}$ satisfying $|\mathcal{I}| = r' \leq r$, parameters $\epsilon, \delta > 0$

1 use Lemma 5 with $\gamma = \epsilon/4r$ to find k_j corresponding to $\mu(I_j)$ for every $j \in [r']$, with probability at least $1 - \delta/2$

2 use Lemma 21 to test all $\binom{r'+1}{2}$ intervals $\mathcal{I}_K$, where K is any interval in $[r']$, with parameters $\eta = \epsilon/100L$, $\gamma = \epsilon/100$, ϵ also for the lemma's "ϵ" parameter, and $\delta/2$

3 let $\mathcal{A}$ denote the set of intervals K for which $\mathcal{I}_K$ was accepted

4 **if** *there exist disjoint* $K_1, \dots, K_s \in \mathcal{A}$ *where* $s \leq L$ *and* $\sum_{j \in [r'] \setminus \bigcup_{t=1}^{s} K_t} 2^{k_j+1} \leq 4r$ **then return** $\mathcal{K} = \left(([r'] \setminus \bigcup_{t=1}^{s} K_t), K_1, \dots, K_s\right)$ **else** `reject`

We now present our departitioning algorithm under the unconditional model, as Algorithm 4.

Let us first prove for the algorithm a soundness lemma, that the probability of outputting a grouping sequence that is not calm is bounded by δ.

Lemma 23 *The probability, for any μ and $\mathcal{I}$, for Algorithm 4 to output (without rejecting) a grouping sequence which is not ϵ-calm is at most δ.*

Proof By a union bound, with probability at least $1-\delta$, both Step 1 produced $k_1, \dots, k_{r'}$ which satisfy $\mu(I_j) \leq \epsilon 2^{k_j+1}/4r$ for every $j \in [r']$, and Step 2 lead to an $\mathcal{A}$ which contains only intervals K for which $d(\mu\restriction_{\mathcal{I}_K}, \mathcal{U}_{\mathcal{I}_K}) \leq \epsilon$. When the two events occur, Step 4 makes sure to output an ϵ-calm grouping sequence (or reject if it fails to find one).

Next, let us prove a completeness lemma. In this case, it depends on both the original distribution being decomposable and the original partition being fine.

Lemma 24 *If μ is $(\epsilon/100, L)$-decomposable, and $\mathcal{I}$ is $(\epsilon/100L, \epsilon/100)$-fine, then with probability at least $1-\delta$, the algorithm will not reject and will output an ϵ-calm grouping sequence.*

Proof As before, with probability at least $1 - \delta$ both Step 1 and Step 2 are successful. As above this event implies that if a sequence is output then it is ϵ-calm (it is the same event as above), so it remains to prove that rejection will not occur. First note that the above event additionally implies that the intervals $K_1, \dots, K_s$ that are provided by Lemma 22 (with $\epsilon/100$ instead of γ) are in particular members of $\mathcal{A}$, by the assertion of Lemma 21.

In particular this means that Step 4 considers the above $K_1, \dots, K_s$. Then, Lemma 5 asserts (due to the event of being successful) that $k_1, \dots, k_{r'}$ are such that for every K', $\sum_{j \in K'} \epsilon 2^{k_j+1}/4r \leq \epsilon/2 + 4\mu(I_{K'})$ (since $\epsilon 2^{k_j+1}/4r \leq \max\{4\mu(I_j), \epsilon/2r\}$). In particular this occurs for $K' = [r'] \setminus \bigcup_{t=1}^{s} K_t$, which by Lemma 22 satisfies $\mu(\mathcal{I}_{K'}) \leq \epsilon/20$, and hence Step 4 will not reject, as it has at least one viable output option.

Finally, let us bound the sample complexity.

Lemma 25 *Algorithm 4 takes* $O(r\log(r/\delta)/\epsilon + \sqrt{nL}\log(r/\delta)/\epsilon^3)$ *many unconditional samples.*

Proof This follows respectively from Lemma 5 with the parameters provided in Step 1, and Lemma 21 with the parameters provided in Step 2.

8 Learning and testing decomposable distributions and properties

Here we finally put things together to produce a learning algorithm for L-decomposable distribution. This algorithm is not only guaranteed to learn with high probability a distribution that is decomposable, but is also guaranteed with high probability to not produce a wrong output for any distribution (though it may plainly reject a distribution that is not decomposable).

This is presented as Algorithm 5. We present it with a fixed error probability 2/3 because this is what we use later on, but it is not hard to move to a general δ.

Algorithm 5: Learning an L-decomposable distribution

Input: Oracle access to a distribution μ supported over $[n]$, parameters L (decomposability), $\epsilon > 0$ (accuracy), a weakly tolerant interval uniformity tester $\mathbb{T}$ taking input values $(\mu, I, m, \gamma, \epsilon, \delta)$

1 use Algorithm 2 with input values $(\mu, \epsilon/2000L, \epsilon/2000, 1/9)$ to obtain a partition $\mathcal{I}$ with $|\mathcal{I}| \le r = 10^5 L\log(1/\epsilon)/\epsilon$
2 use Algorithm 3 with input values $(\mu, L, r, \mathcal{I}, \epsilon/20, 1/9, \mathbb{T})$
3 **if** *Algorithm 3 rejected* **then** `reject`
4 use Lemma 11 with values $(\mu, \mathcal{I}, \epsilon/10, 1/9)$ to obtain $\mu'_{\mathcal{I}}$
5 **return** $\mu'_{\mathcal{I}}$

First we show completeness, that the algorithm will be successful for decomposable distributions.

Lemma 26 *If* μ *is* $(\epsilon/2000, L)$*-decomposable, then with probability at least* $2/3$ *Algorithm 5 produces a distribution* μ' *so that* $d(\mu, \mu') \le \epsilon$.

Proof By Lemma 7, with probability at least 8/9 the partition $\mathcal{I} = (I_1, \ldots, I_r)$ is $(\epsilon/2000L, \epsilon/2000)$-fine, which then implies through use of Lemma 9 that $\sum_{j\in[r]:\mathrm{bias}(\mu\restriction_{I_j})>\epsilon/2000} \mu(I_j) \le 3\epsilon/2000$. When this occurs, by Lemma 17 with probability at least 8/9 Algorithm 3 will accept and so the algorithm will move past Step 3. In this situation, in particular by Lemma 10 we have that $d(\mu_{\mathcal{I}}, \mu) \le 15\epsilon/20$ (in fact this can be bounded much smaller here), and with probability at least 8/9 (by Lemma 11) Step 4 provides a distribution that is $\epsilon/10$-close to $\mu_{\mathcal{I}}$ and hence ϵ-close to μ.

Next we show soundness, that the algorithm will with high probability not mislead about the distribution, whether it is decomposable or not.

Lemma 27 *For any μ, the probability that Algorithm 5 produces (without rejecting) a distribution μ' for which $d(\mu,\mu') > \epsilon$ is bounded by $1/3$.*

Proof Consider the interval partition $\mathcal{I}$. By Lemma 7, with probability at least $8/9$ it is $(\epsilon/2000L, \epsilon/2000)$-fine. When this happens, if $\mathcal{I}$ is such that $\sum_{j:d(\mu\restriction_{I_j},\mathcal{U}_{I_j})>\epsilon/20} \mu(I_j) > 7\epsilon/20$, then by Lemma 18 with probability at least $8/9$ the algorithm will reject in Step 3, and we are done (recall that here a rejection is an allowable outcome).

On the other hand, if $\mathcal{I}$ is such that $\sum_{j:d(\mu\restriction_{I_j},\mathcal{U}_{I_j})>\epsilon/20} \mu(I_j) \leq 7\epsilon/20$, then by Lemma 10 we have that $d(\mu_{\mathcal{I}}, \mu) \leq 15\epsilon/20$, and with probability at least $8/9$ (by Lemma 11) Step 4 provides a distribution that is $\epsilon/10$-close to $\mu_{\mathcal{I}}$ and hence ϵ-close to μ, which is also an allowable outcome.

And finally, we plug in the sample complexity bounds.

Lemma 28 *Algorithm 5 requires $O(L\log(1/\epsilon)/\epsilon + q/\epsilon + L/\epsilon^3)$ many samples, where the value $q = q(n/L, \epsilon^2/10^5 L\log(1/\epsilon), \epsilon/20, \epsilon/500)$ is a bound on the number of samples that each invocation of $\mathbb{T}$ inside Algorithm 3 requires.*

Algorithm 5 can be implemented either as an unconditional sampling algorithm taking $\sqrt{nL}\cdot \mathrm{polylog}(1/\epsilon)/\epsilon^4$ many samples, an adaptive conditional sampling algorithm taking $L\cdot\mathrm{polylog}(1/\epsilon)/\epsilon^3$ many samples, or a non-adaptive conditional sampling algorithm taking $L\cdot \mathrm{polylog}(n)/\mathrm{poly}(\epsilon)$ many samples.

Proof The three summands in the general expression follow respectively from the sample complexity calculations of Lemma 7 for Step 1, Lemma 19 for Step 2, and Lemma 11 for Step 4 respectively. Also note that all samples outside Step 2 are unconditional.

The specific bounds for the two conditional sampling models follow from the respective bounds stated in Lemma 19, while for the unconditional model we replace the middle summand with the bound provided by Lemma 20 (the implementation using parallel testing).

Let us now summarize the above as a theorem.

Theorem 1 *Algorithm 5 is capable of learning an $(\epsilon/2000, L)$-decomposable distribution, giving with probability at least $2/3$ a distribution that is ϵ-close to it, while for no distribution will it give as output a distribution ϵ-far from it with probability more than $1/3$.*

It can be implemented either as an unconditional sampling algorithm taking $\sqrt{nL}\cdot \mathrm{polylog}(1/\epsilon)/\epsilon^4$ many samples, an adaptive conditional sampling algorithm taking $L\cdot \mathrm{polylog}(1/\epsilon)/\epsilon^3$ many samples, or a non-adaptive conditional sampling algorithm taking $L\cdot \mathrm{polylog}(n)/\mathrm{poly}(\epsilon)$ many samples.

Proof This follows from Lemmas 26, 27 and 28 respectively.

For the model of unconditional samples, there is an alternative to Algorithm 5, which uses the departitioning Algorithm 4 instead of the assessment Algorithm 3. We present it as Algorithm 6.

Also for this alternative we will prove completeness, soundness and a sample bound. Completeness is done in a similar manner as Lemma 26.

Algorithm 6: Learning an L-decomposable distribution – departitioning alternative

Input: Oracle access to a distribution μ supported over $[n]$, parameters L (decomposability) and $\epsilon > 0$ (accuracy)

1. use Algorithm 2 with input values $(\mu, \epsilon/2000L, \epsilon/2000, 1/9)$ to obtain a partition $\mathcal{I}$ with $|\mathcal{I}| \leq r = 10^5 L \log(1/\epsilon)/\epsilon$
2. use Algorithm 4 with input values $(\mu, L, r, \mathcal{I}, \epsilon/20, 1/9)$
3. **if** *Algorithm 4 rejected* **then** **`reject`** **else** let $\mathcal{K}$ be its output
4. use Lemma 11 with values $(\mu, \mathcal{I}_{\mathcal{K}}, \epsilon/10, 1/9)$ to obtain $\mu'_{\mathcal{I}_{\mathcal{K}}}$
5. **return** $\mu'_{\mathcal{I}_{\mathcal{K}}}$

Lemma 29 *If μ is $(\epsilon/2000, L)$-decomposable, then with probability at least $2/3$ Algorithm 6 produces a distribution μ' so that $d(\mu, \mu') \leq \epsilon$.*

Proof By Lemma 7, with probability at least 8/9 the obtained partition $\mathcal{I}$ is $(\epsilon/2000L, \epsilon/2000)$-fine, which means that by Lemma 24, with probability at least 8/9 Algorithm 4 will accept and provide a $\mathcal{K}$ which is $\epsilon/20$-calm. In this situation, in particular by Observation 5 we have that $d(\mu_{\mathcal{I}_{\mathcal{K}}}, \mu) \leq 3\epsilon/20$ and with probability at least 8/9 (by Lemma 11) Step 4 provides a distribution that is $\epsilon/10$-close to $\mu_{\mathcal{I}_{\mathcal{K}}}$, and in particular ϵ-close to μ.

Soundness is even easier, because most of the work was already done by Lemma 23.

Lemma 30 *For any μ, the probability that Algorithm 6 produces (without rejecting) a distribution μ' for which $d(\mu, \mu') > \epsilon$ is bounded by $1/3$.*

Proof By Lemma 23, with probability at least 8/9 the invocation in Step 2 will not produce a grouping sequence $\mathcal{K}$ which is not $\epsilon/10$-calm. If there was no rejection in Step 3 (which would have been also allowable), then we conclude using Observation 5 and Lemma 11, exactly as in the proof of Lemma 29 above, that the output distribution is ϵ-close to μ.

We conclude with the sample complexity calculation and the summarizing theorem.

Lemma 31 *Algorithm 6 requires $\sqrt{nL}\log(L) \cdot \mathrm{polylog}(1/\epsilon)/\epsilon^3$ many unconditional samples.*

Proof This follows from plugging in the parameter values of Step 2 in Lemma 25. The sample complexity in Step 4 is subsumed in this.

Theorem 2 *Algorithm 6, using $\sqrt{nL}\log(L) \cdot \mathrm{polylog}(1/\epsilon)/\epsilon^3$ many unconditional samples, is capable of learning an $(\epsilon/2000, L)$-decomposable distribution, giving with probability at least $2/3$ a distribution that is ϵ-close to it, while for no distribution will it give as output a distribution ϵ-far from it with probability more than $1/3$.*

Proof This follows from Lemmas 31, 29 and 30 respectively.

Let us finally move to the immediate application of the above learning algorithms for testing decomposable properties. The algorithm achieving this is presented as Algorithm 7

Algorithm 7: Testing L-decomposable properties.

Input: Oracle access to a distribution μ supported over $[n]$, function $L : (0, 1] \times \mathbb{N} \to \mathbb{N}$ (decomposability), parameter $\epsilon > 0$ (accuracy), an L-decomposable property $\mathcal{C}$ of distributions, a weakly tolerant interval uniformity tester $\mathbb{T}$ taking input values $(\mu, I, m, \gamma, \epsilon, \delta)$

1 use either Algorithm 5 with input values $(\mu, L(\epsilon/4000, n), \epsilon/2, \mathbb{T})$ (all models), or Algorithm 6 with input values $(\mu, L(\epsilon/4000, n), \epsilon/2)$ (unconditional model only), to obtain μ'

2 **if** *Algorithm 5 or 6 accepted and μ' is $\epsilon/2$-close to $\mathcal{C}$* **then** `accept` **else** `reject`

Theorem 3 *Algorithm 7 is a test (with error probability* $1/3$*) for the L-decomposable property $\mathcal{C}$. For $L = L(\epsilon/4000, n)$, It can be implemented either as an unconditional sampling algorithm taking either* $\sqrt{nL} \cdot \text{polylog}(1/\epsilon)/\epsilon^4$ *or* $\sqrt{nL}\log(L) \cdot \text{polylog}(1/\epsilon)/\epsilon^3$ *many samples, an adaptive conditional sampling algorithm taking* $L \cdot \text{polylog}(1/\epsilon)/\epsilon^3$ *many samples, or a non-adaptive conditional sampling algorithm taking* $L \cdot \text{polylog}(n)/\text{poly}(\epsilon)$ *many samples.*

Proof The number and the nature of the samples are determined fully by the application of Algorithm 5 or Algorithm 6 in Step 1, and so follow from either Theorem 1 or Theorem 2 respectively. Also by these theorems, for a distribution $\mu \in \mathcal{C}$, with probability at least $2/3$ an $\epsilon/2$-close distribution μ' will be produced, and so it will be accepted in Step 2.

Finally, if μ is ϵ-far from $\mathcal{C}$, then with probability at least $2/3$ Step 1 will either produce a rejection, or again produce μ' that is $\epsilon/2$-close to μ. In the latter case, μ' will be $\epsilon/2$-far from $\mathcal{C}$ by the triangle inequality, and so Step 2 will reject in both cases.

9 A weakly tolerant tester for the non-adaptive conditional model

Given a distribution μ, supported over $[n]$, and an interval $I \subseteq [n]$ such that $\mu(I) \geq \gamma$, we give a tester that uses non-adaptive conditional queries to μ to distinguish between the cases $\text{bias}(\mu \upharpoonright_I) \leq \epsilon/100$ and $d(\mu \upharpoonright_I, \mathcal{U}_I) > \epsilon$, using ideas from [9]. A formal description of the test is given as Algorithm 8. It is formulated here with error probability $\delta = 1/3$. Lemma 15 is obtained from this the usual way, by repeating the algorithm $O(1/\delta)$ times and taking the majority vote.

We first state the observation that makes Algorithm 8 suitable for a non-adaptive setting.

Observation 6 *Algorithm 8 can be implemented using only* $\text{polylog}(n)/\gamma \cdot \text{poly}(\epsilon)$ *many non-adaptive conditional queries to the distribution μ, that are chosen independently of I.*

Algorithm 8: Non-adaptive weakly tolerant uniformity tester

Input: Oracle access to a distribution μ supported over $[n]$, interval $I \subseteq [n]$, weight bound γ, accuracy $\epsilon > 0$

1 sample $t = \frac{4(\log^{10} n+4)}{\epsilon^2\gamma}$ elements from μ
2 **for** $k \in \{0, \dots, \log n\}$ **do**
3 set $p_k = 2^{-k}$
4 choose a set $U_k \subseteq [n]$, where each $i \in [n]$ is in U_k with probability p_k, independently of other elements in $[n]$
5 **if** $|I| \leq \log^{10} n$ **then**
6 use Lemma 32, using the t unconditional samples from μ, to construct a distribution μ'
7 **if** $d(\mu', \mathcal{U}_I) \leq \epsilon/2$ **then** `accept` **else** `reject`
8 **else**
9 **for** U_k *such that* $k \leq \log\left(\frac{|I|}{2\log^8 n}\right)$*, and* $|I \cap U_k| \geq \log^8 n$ **do**
10 use Lemma 33 for U_k and I, with parameters $\gamma/100\log^2 n$ and $\delta = 1/10\log n$, to find a corresponding η_k, using $\mathrm{polylog}(n)/\gamma$ many samples
11 **if** $\eta_k > \gamma/100\log^2 n$ **then**
12 sample $\log^3 n/\eta_k$ elements from $\mu\restriction_{U_k}$
13 **if** *the same element from* $I \cap U_k$ *has been sampled twice* **then** `reject`
14 choose an index k such that $\frac{2}{3}\log^8 n/\epsilon^2 \leq |I|p_k < \frac{4}{3}\log^8 n/\epsilon^2$
15 sample $m_k = \frac{10^8 \log^{16} n \log(3\log n)}{\epsilon^6\gamma}$ elements from $\mu\restriction_{U_k}$
16 **if** $|I \cap U_k| > 2|I|p_k$ *or the number of samples in* $I \cap U_k$ *is less than* $\gamma m_k/40$ **then**
17 `reject`
18 **else**
19 use Lemma 4 with the samples received from $I \cap U_k$, to construct μ', supported on $I \cap U_k$, such that $\|\mu' - \mu\restriction_{I\cap U_k}\|_\infty \leq \frac{\epsilon}{80|I\cap U_k|}$ with probability at least $9/10$
20 **if** $\|\mu' - \mathcal{U}_{I\cap U_k}\|_\infty \leq \frac{3\epsilon}{80|I\cap U_k|}$ **then** `accept` **else** `reject`

Proof First, note that the algorithm samples elements from μ in four places. Initially, it samples unconditionally from μ in Step 1, and then it performs conditional samples from the sets U_k in Steps 10, 12 and 15. In Steps 10 and 15, the samples are conditioned on sets U_k, where k depends on I. However, observe that we can sample from all sets U_k, for all $0 \leq k \leq \log n$, at the beginning, and then use just the samples taken from the appropriate U_k at these steps. This only increases the bound on the number of samples by a factor of $\log n$. Step 12 depends also on the result of Step 10. However, this step can be reached only for $\eta_k > \gamma/100\log^2 n$. Therefore, we can also here take in advance $100\log^5 n/\gamma$ samples for every possible k, and then use only as many of these samples as we need for this step, for every k for which it is reached. Thus we have only non-adaptive queries, all of which are made at the start of the algorithm, and in a manner independent of I.

The following lemma is used in Step 6 of our algorithm.

Lemma 32 *Let μ be a distribution supported over $[n]$ and $I \subseteq [n]$ be an interval such that $\mu(I) \geq \gamma$. Using $t = \frac{4(|I|+\log(4/\delta))}{\epsilon^2\gamma}$ unconditional queries to μ, we can construct a distribution μ' over I such that, with probability at least $1-\delta$, $d(\mu\restriction_I, \mu') \leq \epsilon$ (in other cases μ' may be arbitrary).*

Proof Take $t = \frac{4(|I|+\log(4/\delta))}{\epsilon^2\gamma}$ unconditional samples. Let t_I be the number of samples that belong to I. Then, $E[t_I] = t\mu(I) \geq t\gamma$. Therefore, by Lemma 1, with probability at least $1 - \exp(-t\mu(I)/4) > 1 - \delta/2$, $t_I \geq t\mu(I)/2 \geq t\gamma/2$.

The t_I samples are distributed according to $\mu\restriction_I$. By the choice of t, the above implies that with probability at least $1-\delta/2$, $t_I \geq 2(|I| + \log(4/\delta))/\epsilon^2$. Therefore, by Lemma 3, we can obtain a distribution μ', supported over I, such that with probability at least $1-\delta/2$, $d(\mu\restriction_I, \mu') \leq \epsilon$. The probability of both of the above events happening is at least $1-\delta$.

If we did not obtain sufficiently many samples (either because $\mu(I) < \gamma$ or due to a low probability event) then we just output an arbitrary distribution supported on I.

The following very simple lemma is used in Step 10 of our algorithm.

Lemma 33 *Given a set $U \subseteq [n]$, an interval $I \subseteq [n]$ and any $\gamma, \delta > 0$, it takes $O(\log(1/\delta)/\gamma)$ samples from μ conditioned on U to output η satisfying the following.*

- $\mu\restriction_U (I) < 2\eta$.
- *If* $\mu\restriction_U (I) \geq \gamma$ *then* $\mu\restriction_U (I) \geq \eta/2$.

Proof We use here Lemma 5, with $\mu\restriction_U$ instead of μ, and with only one set $A_1 = U \cap I$. We then set $\eta = 2^{k_1}\gamma$.

We now move to prove the completeness and soundness of Algorithm 8. In the following analysis, all "$o(1)$ probability" bounds are functions of n only, so these can be made small enough by assuming that n is larger than some global constant; to handle smaller n a brute-force algorithm can be used instead.

Lemma 34 (Completeness) *If $\mu(I) \geq \gamma$ and $\mathrm{bias}(\mu\restriction_I) \leq \epsilon/100$, then Algorithm 8 accepts with probability at least $2/3$.*

Proof First note that if $|I| \leq \log^{10} n$, then we use Lemma 32 to test the distance of $\mu\restriction_I$ to uniform with probability at least $9/10$ in Step 7. For the remaining part of the proof, we will assume that $|I| > \log^{10} n$.

Now note that with probability at least $9/10$, all η_k which were produced in Step 10 indeed satisfy $\mu\restriction_{U_k} (I) \leq 2\eta_k$, and hence $\mu(U_k \cap I) \leq 2\eta_k\mu(U_k)$.

For a set U_k chosen by the algorithm, and any $i \in I \cap U_k$, the probability that it is sampled twice in Step 12 (if it was reached) is at most $\binom{\log^3 n/\eta_k}{2}\left(\frac{\mu(i)}{\mu(U_k)}\right)^2$. Since $\mu(U_k) \geq \mu(I \cap U_k)/2\eta_k$, the probability of sampling i twice in Step 12 is at most $\binom{\log^3 n/\eta_k}{2}\left(\frac{2\eta_k\mu(i)}{\mu(I\cap U_k)}\right)^2$. By Observation 1

$\text{bias}(\mu\restriction_I) \leq \epsilon/100$ implies $\|\mu\restriction_I - \mathcal{U}_I\|_\infty \leq \frac{\epsilon}{100|I|}$, so we have

$$\frac{\mu(I)}{|I|}\left(1-\frac{\epsilon}{100}\right) \leq \mu(i) \leq \frac{\mu(I)}{|I|}\left(1+\frac{\epsilon}{100}\right). \tag{3}$$

From Equation 3 we get the following for all U_k.

$$\frac{|I\cap U_k|\mu(I)}{|I|}\left(1-\frac{\epsilon}{100}\right) \leq \mu(I\cap U_k) \leq \frac{|I\cap U_k|\mu(I)}{|I|}\left(1+\frac{\epsilon}{100}\right). \tag{4}$$

Therefore, the probability that the algorithm samples the same element in $I\cap U_k$ at Step 13 twice is bounded as follows.

$$\begin{aligned}\sum_{i\in I\cap U_k}\binom{\log^3 n/\eta_k}{2}\left(\frac{2\eta_k\mu(i)}{\mu(I\cap U_k)}\right)^2 &\leq 10|I\cap U_k|\binom{\log^3 n}{2}\frac{\max_{i\in I\cap U_k}\mu(i)^2}{\mu(I\cap U_k)^2}\\ &\leq \frac{10}{|I\cap U_k|}\binom{\log^3 n}{2}\left(\frac{1+\epsilon/100}{1-\epsilon/100}\right)^2\end{aligned}$$

Since $|I\cap U_k| \geq \log^8 n$ for any k that goes through Step 9, we can bound the sum as follows.

$$\sum_{i\in I\cap U_k}\binom{\log^3 n/\eta_k}{2}\left(\frac{\mu(i)}{\mu(I\cap U_k)}\right)^2 \leq \frac{10}{\log^2 n}\left(\frac{1+\epsilon/100}{1-\epsilon/100}\right)^2.$$

Therefore, with probability at least $1-o(1)$, the algorithm does not reject at Step 13 for any k for which it is reached.

To show that the algorithm accepts with probability at least $2/3$ in Step 20 (and reaches it), we proceed as follows. Combining Equations 3 and 4, we get the following.

$$\frac{1}{|I\cap U_k|}\left(\frac{1-\epsilon/100}{1+\epsilon/100}\right) \leq \mu\restriction_{I\cap U_k}(i) \leq \frac{1}{|I\cap U_k|}\left(\frac{1+\epsilon/100}{1-\epsilon/100}\right)$$

From this it follows that $\|\mu\restriction_{I\cap U_k} - \mathcal{U}_{I\cap U_k}\|_\infty \leq \frac{\epsilon}{40|I\cap U_k|}$.

We now argue that in this case, the test does not reject in Step 16, for the k chosen in Step 14. Observe that $E[\mu(I\cap U_k)] \geq p_k\gamma$. Also, the expected size of the set $I\cap U_k$ is $p_k|I|$. Since the k chosen in Step 14 is such that $|I|p_k \geq \frac{2}{3}\log^8 n/\epsilon^2 > \frac{2}{3}\log^8 n$, with probability at least $1-\exp(-\Theta(\log^8 n))$, $p_k|I|/2 \leq |I\cap U_k| \leq 2p_k|I|$ (and in particular Step 16 does not reject by its first condition). Therefore, from Equation 4, we get that with probability at least $1-\exp(-\Theta(\log^8 n))$, $\mu(I\cap U_k) \geq p_k\gamma/3$. Since $E[\mu(U_k)] = p_k$, we can conclude using Markov's inequality that, with probability at least $9/10$, $\mu(U_k) \leq 10p_k$. The expected number of samples from $I\cap U_k$ among the m_k samples taken in Step 15 is $m_k\mu(I\cap U_k)/\mu(U_k)$. Therefore, with probability at least $9/10$, the expected number of samples from $I\cap U_k$ among the m_k samples is at least $\gamma m_k/30$. Therefore, with probability, at least $9/10-o(1)$, at least $\gamma m_k/40$ elements of $I\cap U_k$ are sampled, and the tester does not reject in Step 16 by its second condition. The indexes that are sampled in Step 15 that lie in $I \cap U_k$ are distributed according to $\mu\restriction_{I\cap U_k}$, and we know that $|I \cap U_k| \le 2|I|p_k \le \frac{8}{3}\log^8 n/\epsilon^2$. Therefore, with probability at least 9/10, we get a distribution μ' such that $\|\mu' - \mu\restriction_{I\cap U_k}\|_\infty \le \frac{\epsilon}{80|I\cap U_k|}$ in Step 19, and the test correctly accepts in Step 20 for the k chosen in Step 14.

Now we prove the soundness of the tester mentioned above. First we state a lemma from Chakraborty et al [9].

Lemma 35 ([9], adapted for intervals) *Let μ be a distribution, and $I \subseteq [n]$ be an interval such that $d(\mu\restriction_I, \mathcal{U}_I) \ge \epsilon$. Then the following two assertions hold.*

1. *The set $B_1 = \left\{ i \in I \mid \mu\restriction_I(i) < \frac{1+\epsilon/3}{|I|} \right\}$ is such that $|B_1| \ge \epsilon|I|/2$.*
2. *There exists an index $j \in \left\{3, \ldots, \frac{\log|I|}{\log(1+\epsilon/3)}\right\}$, such that the set set $B_j = \{j \in I \mid \frac{(1+\epsilon/3)^{j-1}}{|I|} \le \mu\restriction_I(i) < \frac{(1+\epsilon/3)^j}{|I|}\}$ is of cardinality at least $\frac{\epsilon^2|I|}{96(1+\epsilon/3)^j \log|I|}$.*

Now we analyze the case where $d(\mu\restriction_I, \mathcal{U}_I) > \epsilon$.

Lemma 36 (Soundness) *Let μ be a distribution supported on $[n]$, and let $I \subseteq [n]$ be an interval such that $\mu(I) \ge \gamma$. If $d(\mu\restriction_I, \mathcal{U}_I) \ge \epsilon$, then Algorithm 8 rejects with probability at least* 2/3.

Proof Observe that when $|I| \le \log^{10} n$, the algorithm rejects with probability at least 9/10 in Step 7. For the remainder of the proof, we will assume that $|I| > \log^{10} n$. We analyze two cases according to the value of j given by Lemma 35.

Suppose first that $j > 2$ is such that $|B_j| \ge \frac{\epsilon^2|I|}{96(1+\epsilon/3)^j \log|I|}$, and $(1 + \epsilon/3)^j \le \log^6 n$. The expected number of elements from this set that is chosen in U_k is at least $\frac{\epsilon^2|I|p_k}{96(1+\epsilon/3)^j \log|I|}$. For the choice of k made in Step 14, we have $|I|p_k \ge \frac{2}{3}\log^8 n/\epsilon^2$. The probability that no index from B_j is chosen in U_k is $(1-p_k)^{|B_j|}$, which is at most $(1 - \frac{2\log^8 n}{3|I|\epsilon^2})^{\epsilon^2|I|/96(1+\epsilon/3)^j \log|I|}$. Since $(1+\epsilon/3)^j \le \log^6 n$, this is at most $\exp\left(-\frac{\log n}{144}\right)$. Therefore, with probability $1 - o(1)$, at least one element i is chosen from B_j.

Since $|B_1| \ge \epsilon|I|/2$, the probability that no element from B_1 is chosen in U_k is at most $(1-p_k)^{\epsilon|I|/2}$. Substituting for p_k, we can conclude that, with probability $1 - o(1)$, at least one element i' is chosen from the set B_1.

Now, $\mu\restriction_I(i) \ge (1+\epsilon/3)\mu\restriction_I(i')$. Hence, $\mu\restriction_{I\cap U_k}(i) \ge (1+\epsilon/3)\mu\restriction_{I\cap U_k}(i')$. This implies that $\|\mu\restriction_{I\cap U_k} - \mathcal{U}_{I\cap U_k}\|_\infty \ge \frac{\epsilon}{20|I\cap U_k|}$. The algorithm will hence reject with probability at least 9/10 in Step 20, unless it has already rejected in Step 16 or Step 13.

Now assume the second case, where j is such that $|B_j| \ge \frac{\epsilon^2|I|}{96(1+\epsilon/3)^j \log|I|}$, and $(1+\epsilon/3)^j > \log^6 n$. Let $k = \max\left\{0, \lfloor \log\left(\frac{\epsilon^2|I|}{4(1+\epsilon/3)^j \log^2 n}\right)\rfloor\right\}$. Then, for this value of k, $p_k \ge \min\left\{1, \frac{2(1+\epsilon/3)^j \log^2 n}{\epsilon^2|I|}\right\}$. Also, for this value of k,

$p_k \le \min\left\{1, \frac{4(1+\epsilon/3)^j \log^2 n}{\epsilon^2|I|}\right\}$. By Lemma 1, with probability at least $1 - \exp(-\Theta(\log^8 n))$, $|U_k \cap I| \ge \log^8 n$, for this value of k. Thus, with probability $1 - o(1)$ Step 9 will allow steps 10 through Step 13 to take place for this k.

Furthermore, the probability that $B_j \cap U_k$ is empty is $(1-p_k)^{|B_j|}$. Substituting the values of $|B_j|$ and p_k, we get that $\Pr[B_j \cap U_k = \emptyset] \le \exp(-\log n/48)$. Therefore, with probability at least $1-\exp(-\log n/48)$, U_k contains an element of B_j.

Let $i \in B_j \cap U_k$. Since $i \in B_j$, from Lemma 35 we know that $\mu\restriction_I(i) = \frac{\mu(i)}{\mu(I)} \ge \frac{(1+\epsilon/3)^{j-1}}{|I|}$. From this bound, we get that $\mu\restriction_{U_k}(i) \ge \frac{(1+\epsilon/3)^{j-1}\mu(I)}{|I|\mu(U_k)}$. The expected value of $\mu(U_k)$ is p_k. By Markov's inequality, with probability at least $9/10$, $\mu(U_k) \le 10p_k$. Therefore, $\mu\restriction_{U_k}(i) \ge \frac{(1+\epsilon/3)^{j-1}\gamma}{10|I|p_k} \ge \frac{\gamma}{40\epsilon^2(1+\epsilon/3)\log^2 n}$. In particular, since $\mu\restriction_{U_k}(I) \ge \mu\restriction_{U_k}(i)$, with probability $1-o(1)$ the η_k that is produced for it in Step 10 is such that Step 11 allows Steps 12 and 13 to take place.

The probability that in Step 12 i is sampled less than twice is the sum of the probability that it is not sampled at all, at most $\left(1 - \frac{\gamma}{40\epsilon^2(1+\epsilon/3)\log^2 n}\right)^{\log^3(n)/\gamma}$, and the probability that this index is sampled exactly once, which is at most $\frac{\log^3(n)}{\gamma}\frac{\gamma}{40\epsilon^2(1+\epsilon/3)\log^2 n}\left(1 - \frac{\gamma}{40\epsilon^2(1+\epsilon/3)\log^2 n}\right)^{(\log^3(n)/\gamma)-1}$. Both are $o(1)$ (as a function of n only, not γ or $\epsilon \le 1$), and therefore with probability $1-o(1)$, i is sampled at least twice, and the tester rejects in Step 13.

Proof (Proof of Lemma 15) Given the input values $(\mu, I, m, \gamma, \epsilon, \delta)$, we iterate Algorithm 8 $O(1/\delta)$ independent times with input values $(\mu, I, \gamma, \epsilon)$ (we may ignore m here), and take the majority vote. The sample complexity is evident from the description of the algorithm. If indeed $\mu(I) \ge \gamma$ then Lemma 34 and Lemma 36 provide that every round gives the correct answer with probability at least $2/3$, making the majority vote correct with probability at least $1-\delta$. The sample complexity and the independence of the requested samples from I are guaranteed by Observation 6.

10 Introducing properties characterized by atlases

In the next sections, we give a testing algorithm for properties characterized by atlases, which we formally define here. We will also show that distributions that are L-decomposable are, in particular, characterized by atlases. First we start with the definition of an inventory.

Definition 14 (inventory) Given an interval $I = [a,b] \subseteq [n]$ and a real-valued function $\nu : [a,b] \to [0,1]$, the *inventory* of ν over $[a,b]$ is the multiset M corresponding to $(\nu(a), \dots, \nu(b))$.

That is, we keep count of the function values over the interval including repetitions, but ignore their order. In particular, for a distribution $\mu = (p_1, \dots, p_n)$

over $[n]$, the inventory of μ over $[a, b]$ is the multiset M corresponding to $(p_a, \ldots, p_b)$.

Definition 15 (atlas) Given a distribution μ over $[n]$, and an interval partition $\mathcal{I} = (I_1, \ldots, I_k)$ of $[n]$, the *atlas* $\mathcal{A}$ of μ over $\mathcal{I}$ is the ordered pair $(\mathcal{I}, \mathcal{M})$, where $\mathcal{M}$ is the sequence of multisets $(M_1, \ldots, M_k)$ so that M_j is the inventory of μ over I_j, for every $j \in [k]$. In this setting, we also say that μ *conforms* to $\mathcal{A}$.

We note that there can be many distributions over $[n]$ whose atlas is the same. We will also denote by an atlas $\mathcal{A}$ any ordered pair $(\mathcal{I}, \mathcal{M})$ where $\mathcal{I}$ is an interval partition of $[n]$ and $\mathcal{M}$ is a sequence of multisets of the same length, so that the total sum of all members of all multisets is 1. It is a simple observation that for every such $\mathcal{A}$ there exists at least one distribution that conforms to it. The *length* of an atlas $|\mathcal{A}|$ is defined as the shared length of its interval partition and sequence of multisets.

We now define what it means for a property to be characterized by atlases, and state our main theorem concerning such properties.

Definition 16 For a function $k : \mathbb{R}^+ \times \mathbb{N} \to \mathbb{N}$, we say that a property of distributions $\mathcal{C}$ is *k-characterized by atlases* if for every $n \in \mathbb{N}$ and every $\epsilon > 0$ we have a set $\mathbb{A}$ of atlases of lengths bounded by $k(\epsilon, n)$, so that every distribution μ over $[n]$ satisfying $\mathcal{C}$ conforms to some $\mathcal{A} \in \mathbb{A}$, while on the other hand no distribution μ over $[n]$ that conforms to any $\mathcal{A} \in \mathbb{A}$ is ϵ-far from satisfying $\mathcal{C}$.

Theorem 4 *If $\mathcal{C}$ is a property of distributions that is k-characterized by atlases, then for any $\epsilon > 0$ there is an adaptive conditional testing algorithm for $\mathcal{C}$ with query complexity $k(\epsilon/5, n) \cdot \mathrm{polylog}(n)/\mathrm{poly}(\epsilon)$, and error probability bound $1/3$.*

10.1 Applications and examples

We first show that L-decomposable properties are in particular characterized by atlases.

Lemma 37 *If $\mathcal{C}$ is a property of distributions that is L-decomposable, then $\mathcal{C}$ is k-characterized by atlases, where $k(\epsilon, n) = L(\epsilon/3, n)$.*

Proof Every distribution $\mu \in \mathcal{C}$ that is supported over $[n]$ defines an atlas in conjunction with the interval partition of the L-decomposition of μ for $L = L(\gamma, n)$. Let $\mathbb{A}$ be the set of all such atlases. We will show that $\mathcal{C}$ is $L(3\gamma, n)$-characterized by $\mathbb{A}$.

Let $\mu \in \mathcal{C}$. Since μ is L-decomposable, μ conforms to the atlas given by the L-decomposition and it is in $\mathbb{A}$ as defined above.

Now suppose that μ conforms to an atlas $\mathcal{A} \in \mathbb{A}$, where $\mathcal{I} = (I_1, \ldots, I_\ell)$ is the sequence of intervals. By the construction of $\mathbb{A}$, there exists a distribution

$\chi \in \mathcal{C}$ that conforms with $\mathcal{A}$. Now, for each $j \in [\ell]$ such that $\mu(I_j) \leq \gamma/L$, we have (noting that $\chi(I_j) = \mu(I_j)$)

$$\sum_{i \in I_j} |\mu(i) - \chi(i)| \leq \sum_{i \in I_j} \mu(i) + \sum_{i \in I_j} \chi(i) \leq 2\mu(I_j) \leq \frac{2\gamma}{\ell}. \tag{5}$$

Noting that μ and χ have the same maximum and minimum over I_j (as they have the same inventory), for each $j \in [\ell]$ and $i \in I_j$, we know that $|\mu(i) - \chi(i)| \leq \max_{i \in I_j} \mu(i) - \min_{i \in I_j} \mu(i)$. Therefore, for all $j \in [\ell]$ such that $\max_{i \in I_j} \mu(i) \leq (1+\gamma) \min_{i \in I_j} \mu(i)$, $|\mu(i) - \chi(i)| \leq \gamma \min_{i \in I_j} \mu(i)$. Therefore,

$$\sum_{i \in I_j} |\mu(i) - \chi(i)| \leq |I_j| \gamma \min_{i \in I_j} \mu_j(i) \leq \gamma \mu_j(I_j). \tag{6}$$

Finally, recall that since $\mathcal{A}$ came from an L-decomposition of χ, all intervals are covered by the above cases. Summing up Equations 5 and 6 for all $j \in [\ell]$, we obtain $d(\mu, \chi) \leq 3\gamma$.

Note that atlases characterize also properties that do not have shape restriction. The following is a simple observation.

Observation 7 *If $\mathcal{C}$ is a property of distributions that is symmetric over $[n]$, then $\mathcal{C}$ is $\mathbf{1}$-characterized by atlases.*

It was shown in Chakraborty et al [9] that such properties are efficiently testable using conditional queries, so Theorem 4 in particular generalizes this result.

Finally, the notion of characterization by atlases provides a natural model for tolerant testing, as we will see in the next section.

11 Atlas characterizations and tolerant Testing

We now show that for all properties of distributions that are characterized by atlases, there are efficient *tolerant testers* as well. In [8], it was shown that for a large property of distribution properties that have "semi-agnostic" learners, there are efficient tolerant testers. In this section, we show that when the algorithm is given conditional query access, there are efficient tolerant testers for the larger class of properties that are characterized by atlases, including decomposable properties that otherwise do not lend themselves to tolerant testing. The mechanism presented here will also be used in the proof of Theorem 4 itself.

First, we give a definition of tolerant testing. We note that the definition extends naturally to algorithms that make conditional queries to a distribution.

Definition 17 Let $\mathcal{C}$ be any property of probability distributions. An (η, ϵ)-*tolerant tester* for $\mathcal{C}$ with query complexity q and error probability δ, is an algorithm that samples q elements $x_1, \ldots, x_q$ from a distribution μ, accepts with probability at least $1 - \delta$ if $d(\mu, \mathcal{C}) \leq \eta$, and rejects with probability at least $1 - \delta$ if $d(\mu, \mathcal{C}) \geq \eta + \epsilon$.

In [8], they show that for every $\alpha > 0$, there is an $\epsilon > 0$ that depends on α, such that there is an $(\epsilon, \alpha - \epsilon)$-tolerant tester for certain shape-restricted properties. On the other hand, tolerant testing using unconditional queries for other properties, such as the (**1**-decomposable) property of being uniform, requires $\Omega(n/\log n)$ many samples by [19]. We prove that, in the presence of conditional query access, there is an (η, ϵ)-tolerant tester for every $\eta, \epsilon > 0$ such that $\eta + \epsilon < 1$, for all properties of probability distributions that are characterized by atlases.

We first present a definition and prove an easy lemma that will be useful later on.

Definition 18 Given a partition $\mathcal{I} = (I_1, \ldots, I_k)$ of $[n]$, we say that a permutation $\sigma : [n] \to [n]$ is *$\mathcal{I}$-preserving* if for every $1 \le j \le k$ we have $\sigma(I_j) = I_j$.

Lemma 38 *Let χ and χ' be two distributions, supported on $[n]$, both of which conform to an atlas $\mathcal{A} = (\mathcal{I}, \mathcal{M})$. If $\mathcal{A}' = (\mathcal{I}, \mathcal{M}')$ is another atlas with the same interval partition as $\mathcal{A}$, such that χ is ϵ-close to conforming to $\mathcal{A}'$, then χ' is also ϵ-close to conforming to $\mathcal{A}'$.*

Proof It is an easy observation that there exists an $\mathcal{I}$-preserving permutation σ that moves χ to χ'. Now let μ be the distribution that conforms to $\mathcal{A}'$ such that $d(\mu, \chi) \le \epsilon$, and let μ' be the distribution that results from having σ operate on μ. It is not hard to see that μ' conforms to $\mathcal{A}'$ (which has the same interval partition as $\mathcal{A}$), and that it is ϵ-close to χ'.

For a property $\mathcal{C}$ of distributions that is k-characterized by atlases, let $\mathcal{C}_\eta$ be the property of distributions of being η-close to $\mathcal{C}$. The following lemma states that $\mathcal{C}_\eta$ is also k-characterized by atlases. This lemma will also be important for us outside the context of tolerant testing per-se.

Lemma 39 *Let $\mathcal{C}$ be a property of distributions that is k-characterized by atlases. For any $\eta > 0$, let $\mathcal{C}_\eta$ be the set of all probability distributions μ such that $d(\mu, \mathcal{C}) \le \eta$. Then, for every ϵ and n, there is a set $\mathbb{A}_\eta$ of atlases of length at most $k(\epsilon, n)$, such that every $\mu \in \mathcal{C}_\eta$ over $[n]$ conforms to at least one atlas in $\mathbb{A}_\eta$, and every distribution that conforms to an atlas in $\mathbb{A}_\eta$ is $\eta + \epsilon$-close to $\mathcal{C}$, and is ϵ-close to $\mathcal{C}_\eta$.*

Proof Since $\mathcal{C}$ is k-characterized by atlases, for every ϵ and n there is a set of atlases $\mathbb{A}$ of length at most $k(\epsilon, n)$, such that for each $\mu \in \mathcal{C}$ over $[n]$ there is an atlas $\mathcal{A} \in \mathbb{A}$ to which it conforms, and any χ that conforms to an atlas $\mathcal{A} \in \mathbb{A}$ is ϵ-close to $\mathcal{C}$. Now, let $\mathbb{A}_\eta$ be obtained by taking each atlas $\mathcal{A} \in \mathbb{A}$, and adding all atlases, with the same interval partition, corresponding to distributions that are η-close to conforming to $\mathcal{A}$.

First, note that the new atlases that are added have the same interval partitions as atlases in $\mathbb{A}$, and hence have the same length bound $k(\epsilon, n)$. To complete the proof of the lemma, we need to prove that every $\mu \in \mathcal{C}_\eta$ conforms to some atlas in $\mathbb{A}_\eta$, and that no distribution that conforms to $\mathbb{A}_\eta$ is $\eta + \epsilon$-far from $\mathcal{C}$ (and ϵ-far from C_η).

Take any $\mu \in \mathcal{C}_\eta$. There exists some distribution $\mu' \in \mathcal{C}$ such that $d(\mu, \mu') \leq \eta$. Since $\mathcal{C}$ is k-characterized by atlases, there is some atlas $\mathcal{A} \in \mathbb{A}$ such that μ' conforms with $\mathcal{A}$. Also, observe that μ is η-close to conforming to $\mathcal{A}$ through μ'. Therefore, there is an atlas $\mathcal{A}'$ with the same interval partition as $\mathcal{A}$ that was added in $\mathbb{A}_\eta$, which is the atlas corresponding to the distribution μ. Hence, there is an atlas in $\mathbb{A}_\eta$ to which μ conforms.

Conversely, let χ be a distribution that conforms to an atlas $\mathcal{A}' \in \mathbb{A}_\eta$. From the construction of $\mathbb{A}_\eta$, we know that there is an atlas $\mathcal{A} \in \mathbb{A}$ with the same interval partition as $\mathcal{A}'$, and there is a distribution χ' that conforms to $\mathcal{A}'$ and is η-close to conforming to $\mathcal{A}$. Therefore, by Lemma 38 χ is also η-close to conforming to $\mathcal{A}$. Let μ' be the distribution conforming to $\mathcal{A}$ such that $d(\chi, \mu') \leq \eta$. Since μ' conforms to an atlas $\mathcal{A} \in \mathbb{A}$, $d(\mu', \mathcal{C}) \leq \epsilon$. Therefore, by the triangle inequality, $d(\chi, \mathcal{C}) \leq \eta + \epsilon$.

This also implies that $d(\chi, \mathcal{C}_\eta) \leq \epsilon$, by considering $\tilde{\chi} = (\epsilon\tilde{\mu} + \eta\chi)/(\epsilon + \eta)$ where $\tilde{\mu}$ is the distribution in $\mathcal{C}$ that is $\epsilon + \eta$-close to χ. Note that $\tilde{\chi}$ is η-close to $\tilde{\mu}$ and ϵ-close to χ.

Using Lemma 39 we get the following corollary of Theorem 4 about tolerant testing of distributions characterized by atlases.

Corollary 1 *Let $\mathcal{C}$ be a property of distributions that is k-characterized by atlases. For every $\eta, \epsilon > 0$ such that $\eta + \epsilon < 1$, there is an (η, ϵ)-tolerant tester for $\mathcal{C}$ that takes $k(\epsilon/5, n) \cdot \mathrm{poly}(\log n, 1/\epsilon)$ conditional samples and succeeds with probability at least $2/3$.*

12 Some useful lemmas about atlases and characterizations

We start with a definition and a lemma, providing an alternative equivalent definition of properties k-characterizable by atlases

Definition 19 (permutation-resistant distributions) For a function $k : \mathbb{R}^+ \times \mathbb{N} \to \mathbb{N}$, a property $\mathcal{C}$ of probability distributions is *k-piecewise permutation resistant* if for every $n \in \mathbb{N}$, every $\epsilon > 0$, and every distribution μ over $[n]$ in $\mathcal{C}$, there exists a partition $\mathcal{I}$ of $[n]$ into up to $k(\epsilon, n)$ intervals, so that every $\mathcal{I}$-preserving permutation of $[n]$ transforms μ into a distribution that is ϵ-close to a distribution in $\mathcal{C}$.

Lemma 40 *For $k : \mathbb{R}^+ \times \mathbb{N} \to \mathbb{N}$, a property $\mathcal{C}$ of probability distributions is k-piecewise permutation resistant if and only if it is k-characterized by atlases.*

Proof If $\mathcal{C}$ is k-piecewise permutation resistant, then for each distribution $\mu \in \mathcal{C}$, there exists an interval partition $\mathcal{I}_\mu$ of $[n]$, such that every $\mathcal{I}_\mu$-preserving permutation of $[n]$ transforms μ into a distribution that is ϵ-close to $\mathcal{C}$. Each distribution μ thus gives an atlas over $\mathcal{I}_\mu$, and the collection of these atlases for all $\mu \in \mathcal{C}$ characterizes the property $\mathcal{C}$. Therefore, $\mathcal{C}$ is k-characterized by atlases.

Conversely, let $\mathcal{C}$ be a property of distributions that are k-characterized by atlases and let $\mathbb{A}$ be the set of atlases. For each $\mu \in C$, let $\mathcal{A}_\mu$ be the atlas in $\mathbb{A}$ that characterizes μ and let $\mathcal{I}_\mu$ be the interval partition corresponding to this atlas. Now, every $\mathcal{I}_\mu$-preserving permutation σ of μ gives a distribution μ_σ that has the same atlas $\mathcal{A}_\mu$. Since $\mathcal{C}$ is k-characterized by atlases, μ_σ is ϵ-close to $\mathcal{C}$. Therefore, $\mathcal{C}$ is k-piecewise permutation resistant as well.

We now prove the following lemma about ϵ/k-fine partitions of distributions characterized by atlases, having a similar flavor as Lemma 8 for L-decomposable properties. Since we cannot avoid a polylog(n) dependency anyway, for simplicity we use the one-parameter variant of fine partitions.

Lemma 41 *Let $\mathcal{C}$ be a property of distributions that is k-characterized by atlases, and for some ϵ and n let $\mathbb{A}$ be the corresponding characterization. For any $\mu \in \mathcal{C}$, any ϵ/k-fine interval partition $\mathcal{I}'$ of μ, and the corresponding atlas $\mathcal{A}' = (\mathcal{I}', \mathcal{M}')$ for μ (not necessarily in $\mathbb{A}$), any distribution μ' that conforms to $\mathcal{A}'$ is 3ϵ-close to $\mathcal{C}$.*

Proof Let $\mathcal{A} = (\mathcal{I}, \mathcal{M})$ be the atlas from $\mathbb{A}$ to which μ conforms, and let $\mathcal{I}' = (I'_1, I'_2, \ldots, I'_r)$ be an ϵ/k-fine interval partition of μ. Let $\mathcal{P} \subseteq \mathcal{I}'$ be the set of intervals that intersect more than one interval in $\mathcal{I}$. Since $\mathcal{I}'$ is ϵ/k-fine, and the length of $\mathcal{A}$ is at most k, $\mu(\bigcup_{I'_j \in \mathcal{P}} I'_j) \le \epsilon$ (note that $\mathcal{P}$ cannot contain singletons). Also, since μ' conforms to $\mathcal{A}'$, we have $\mu'(\bigcup_{I'_j \in \mathcal{P}} I'_j) \le \epsilon$.

Let $\tilde{\mu}$ be a distribution supported over $[n]$ obtained as follows: For each interval $I'_j \in \mathcal{P}$, $\tilde{\mu}(i) = \mu(i)$ for every $i \in I'_j$. For each interval $I'_j \in \mathcal{I}' \setminus \mathcal{P}$, $\tilde{\mu}(i) = \mu'(i)$ for every $i \in I'_j$. Note that the inventories of $\tilde{\mu}$ and μ are identical over any I'_j in $\mathcal{I}' \setminus \mathcal{P}$. From this it follows that $\tilde{\mu}$ also conforms to $\mathcal{A}$, and in particular $\tilde{\mu}$ is a distribution. To see this, for any I_j in $\mathcal{I}$ partition it to its intersection with the members of $\mathcal{I}' \setminus \mathcal{P}$ contained in it, and all the rest. For the former we use that μ and μ' have the same inventories, and for the latter we specified that $\tilde{\mu}$ has the same values as μ.

Since μ' and $\tilde{\mu}$ are identical at all points except those in $\mathcal{P}$, we have $d(\mu', \tilde{\mu}) \le 2\epsilon$. Furthermore, $d(\tilde{\mu}, \mathcal{C}) \le \epsilon$ since $\tilde{\mu}$ conforms to $\mathcal{A} \in \mathbb{A}$. Therefore, by the triangle inequality, $d(\mu', \mathcal{C}) \le 3\epsilon$.

The main idea of our test, for a property of distributions k-characterized by atlases, starts with a fine partition $\mathcal{I}$ obtained through Algorithm 1. We then show how to compute an atlas $\mathcal{A}$ with this interval partition, such that there is a distribution $\mu_{\mathcal{I}}$ that conforms to $\mathcal{A}$ that is close to μ. We use the trimming sampler from [9] to obtain such an atlas corresponding to $\mathcal{I}$. To test if μ is in $\mathcal{C}$, we show that it is sufficient to check if there is some distribution conforming to $\mathcal{A}$ that is close to a distribution in $\mathcal{C}$.

13 An adaptive test for properties characterized by atlases

Our main technical lemma, which we state here and prove in Section 14, is the following possibility of "learning" an atlas of an unknown distribution for an interval partition $\mathcal{I}$, under the conditional sampling model.

Lemma 42 *Given a distribution μ supported over $[n]$, and a partition $\mathcal{I} = (I_1, I_2, \ldots, I_r)$, using $r \cdot \mathrm{poly}(\log n, 1/\epsilon, \log(1/\delta))$ conditional samples from μ we can construct, with probability at least $1-\delta$, an atlas for some distribution $\mu_{\mathcal{I}}$ that is ϵ-close to μ.*

First, we show how this implies Theorem 4. To prove it, we give as Algorithm 9 a formal description of the test.

Algorithm 9: Adaptive conditional tester for properties k-characterized by atlases

Input: Oracle access to a distribution μ supported over $[n]$, a function $k : (0,1] \times \mathbb{N} \to \mathbb{N}$, accuracy parameter $\epsilon > 0$, a property $\mathcal{C}$ of distributions that is k-characterized by the set of atlases $\mathbb{A}$

1 use Algorithm 1 with input values $(\mu, \epsilon/5k(\epsilon/5, n), 1/6)$ to obtain a partition $\mathcal{I}$ with $|\mathcal{I}| \leq 1 + 30k(\epsilon/5, n)\log(90k(\epsilon/5, n)/\epsilon)/\epsilon = O(k(\epsilon/5, n)\log(n)\log(1/\epsilon)/\epsilon)$

2 use Lemma 42 with accuracy parameter $\epsilon/5$ and error parameter $1/6$ to obtain an atlas $\mathcal{A}_{\mathcal{I}}$ corresponding to $\mathcal{I}$

3 **if** *there exists $\chi \in \mathcal{C}$ that is $\epsilon/5$-close to conforming to $\mathcal{A}_{\mathcal{I}}$* **then** `accept` **else** `reject`

Lemma 43 (completeness) *Let $\mathcal{C}$ be a property of distributions that is k-characterized by atlases, and let μ be any distribution supported over $[n]$. If $\mu \in \mathcal{C}$, then with probability at least $2/3$ Algorithm 9 accepts.*

Proof In Step 2, with probability at least $5/6 > 2/3$, we get an atlas $\mathcal{A}_{\mathcal{I}'}$, such that there is a distribution $\mu_{\mathcal{I}'}$ conforming to it that is $\epsilon/5$-close to μ. Step 3 then accepts on account of $\chi = \mu$.

Lemma 44 (soundness) *Let $\mathcal{C}$ be a property of distributions k-characterized by atlases, and let μ be any distribution supported over $[n]$. If $d(\mu, \mathcal{C}) > \epsilon$, then with probability at least $2/3$ Algorithm 9 rejects.*

Proof With probability at least $2/3$, for $k = k(\epsilon/5, n)$ we get an $\epsilon/5k$-fine partition $\mathcal{I}'$ in Step 1, as well as an atlas $\mathcal{A}_{\mathcal{I}'}$ in Step 2 such that there is a distribution $\mu_{\mathcal{I}'}$ conforming to it that is $\epsilon/5$-close to μ.

Suppose that the algorithm accepted in Step 3 on account of $\chi \in \mathcal{C}$. Then there is a χ' that is $\epsilon/5$-close to χ and conforms to $\mathcal{A}_{\mathcal{I}'}$. By Lemma 39, the property $\mathcal{C}_{\epsilon/5}$ of being $\epsilon/5$-close to $\mathcal{C}$ is itself k-characterized by atlases. Let $\mathbb{A}_{\epsilon/5}$ be the collection of atlases characterizing it. Using Lemma 41 with χ' and $\mathbb{A}_{\epsilon/5}$, we know that χ' is $3\epsilon/5$-close to some χ, which is in $\mathcal{C}_{\epsilon/5}$ and thus $\epsilon/5$-close to $\mathcal{C}$. Since χ' is also $\epsilon/5$-close to μ, we obtain that μ is ϵ-close to $\mathcal{C}$ by the triangle inequality, contradicting $d(\mu, \mathcal{C}) > \epsilon$.

Proof (Proof of Theorem 4) Given a distribution μ, supported on $[n]$, and a property $\mathcal{C}$ of distributions that is k-characterized by atlases, we use Algorithm 9. The correctness follows from Lemmas 43 and 44. The number of samples made in Step 1 is clearly dominated by the number of samples in Step 2, which is $k(\epsilon/5, n) \cdot \mathrm{polylog}(n)/\mathrm{poly}(\epsilon)$.

14 Constructing an atlas for a distribution

Before we prove Lemma 42, we will define the notion of value-distances and prove lemmas that will be useful for the proof of the theorem.

Definition 20 (value-distance) Given two multisets A,B of real numbers, both of the same size (e.g. two inventories over an interval $[a, b]$), the *value-distance* between them is the minimum ℓ_1 distance between a vector that conforms to A and a vector that conforms to B.

The following observation gives a simple method to calculate the value-distances between two multisets A and B.

Observation 8 *The value-distance between A and B is equal to the ℓ_1 distance between the two vectors resulting from the respective sorting of the two multisets.*

Proof Given two vectors v and w corresponding to A and B achieving the value-distance, first assume that the smallest value of A is not larger than the smallest value of B. Assume without loss of generality (by permuting both v and w) that v_1 is of the smallest value among those of A. It is not hard to see that in this case one can make w_1 to be of the smallest value among those of B without increasing the distance (if w_k holds the smallest value, swapping it with w_1 will not increase the distance), and from here one can proceed by induction over $|A| = |B|$.

We now prove two lemmas that will be useful for the proof of Lemma 42.

Lemma 45 *Let A and B be two multisets of the same size, both with members whose values range in $\{0, \alpha_1, \ldots, \alpha_r\}$. Let m_j be the number of appearances of α_j in A, and n_j the corresponding number in B. If $m_j \le n_j$ for every $1 \le j \le r$, then the value-distance between A and B is bounded by $\sum_{j=1}^{r}(n_j - m_j)\alpha_j$.*

Proof Let $v_A = \{a_1, \ldots, a_l\}$ and $v_B = \{b_1, \ldots, b_l\}$ be two vectors such that $a_1 = \cdots = a_{m_1} = \alpha_1$, $a_{m_1+1} = \cdots = a_{n_1} = 0$ and $b_1 = \cdots = b_{n_1} = \alpha_1$, and similarly for $j \in \{1, \ldots, r-1\}$, $a_{\sum_{i=1}^{j} n_i+1} = \cdots = a_{\sum_{i=1}^{j} n_i+m_{j+1}} = \alpha_{j+1}$, $a_{\sum_{i=1}^{j} n_i+m_{j+1}+1} = \cdots = a_{\sum_{i=1}^{j+1} n_i} = 0$ and $b_{\sum_{i=1}^{j} n_i+1} = \cdots = b_{\sum_{i=1}^{j+1} n_i} = \alpha_{j+1}$. For $k > \sum_{j=1}^{r} n_j$, we set $a_k = b_k = 0$. The vectors v_A and v_B conform to the multisets A and B respectively, and the ℓ_1 distance between the two vectors is $\sum_{j=1}^{r}(n_j - m_j)\alpha_j$, so the lemma follows.

Lemma 46 *Let μ be a probability distribution over $\{1,\ldots,n\}$, and let $\tilde{\mu}$ be a vector of size n where each entry is a real number in the interval $[0,1]$, such that $\sum_{i\in[n]}|\mu(i)-\tilde{\mu}(i)|\leq\epsilon$. Let $\hat{\mu}$ be a probability distribution over $\{1,\ldots,n\}$ defined as $\hat{\mu}(i)=\tilde{\mu}(i)/\sum_{i\in[n]}\tilde{\mu}(i)$ for all $i\in[n]$. Then $\sum_{i\in[n]}|\mu(i)-\hat{\mu}(i)|\leq 5\epsilon$.*

Proof We have $|\sum_{i\in[n]}\mu(i)-\sum_{i\in[n]}\tilde{\mu}(i)|\leq\epsilon$. Therefore, $1-\epsilon\leq\sum_{i\in[n]}\tilde{\mu}(i)\leq 1+\epsilon$. If $\sum_{i\in[n]}\tilde{\mu}(i)<1$, then $\hat{\mu}(i)\leq\tilde{\mu}(i)/(1-\epsilon)$ and $\hat{\mu}(i)>\tilde{\mu}(i)$. Therefore, $\hat{\mu}(i)\leq(1+2\epsilon)\tilde{\mu}(i)$ and hence $0\leq\hat{\mu}(i)-\tilde{\mu}(i)\leq 2\epsilon\tilde{\mu}(i)$. If $\sum_{i\in[n]}\tilde{\mu}(i)\geq 1$, then $\hat{\mu}(i)\geq\tilde{\mu}(i)/(1+\epsilon)\geq(1-\epsilon)\tilde{\mu}(i)$ and $\hat{\mu}(i)\leq\tilde{\mu}(i)$. Therefore $0\leq\tilde{\mu}(i)-\hat{\mu}(i)\leq\epsilon\tilde{\mu}(i)$. Therefore $|\tilde{\mu}(i)-\hat{\mu}(i)|\leq 2\epsilon\tilde{\mu}(i)$, in all cases, for all i.

Now, $\sum_{i\in[n]}|\mu(i)-\hat{\mu}(i)|\leq\sum_{i\in[n]}|\mu(i)-\tilde{\mu}(i)|+\sum_{i\in[n]}|\tilde{\mu}(i)-\hat{\mu}(i)|$. Since $\sum_{i\in[n]}|\tilde{\mu}(i)-\hat{\mu}(i)|\leq 2\epsilon\sum_{i\in[n]}\tilde{\mu}(i)\leq 2\epsilon(1+\epsilon)$, we get that $\sum_{i\in[n]}|\mu(i)-\hat{\mu}(i)|\leq 5\epsilon$.

We now recall the definition of an ϵ-trimming sampler from [9].

Definition 21 (ϵ-trimming sampler) An *ϵ-trimming sampler* providing s samples for a distribution μ supported over $[n]$, is an algorithm that has conditional query access to the distribution μ and returns s pairs of values $(r,\bar{\mu}(r))$ (for $r=0$, $\bar{\mu}(r)$ is not output) from a distribution $\bar{\mu}$ supported on $\{0\}\cup[n]$, such that $\sum_{i\in[n]}|\bar{\mu}(i)-\mu(i)|\leq 4\epsilon$, and each r is independently drawn from $\bar{\mu}$. Furthermore, there is a set P of $\mathrm{poly}(\log n,1/\epsilon)$ real numbers such that for all i either $\bar{\mu}(i)=0$ or $\bar{\mu}(i)\in P$.

The existence of an ϵ-trimming sampler with a small set of values was proved in [9]. Let us formally state this lemma.

Lemma 47 ([9]) *Given conditional query access to a distribution μ supported on $[n]$, there is an ϵ-trimming sampler that makes $32s\cdot\epsilon^{-4}\cdot\log^5 n\cdot\log(s\delta^{-1}\log n)$ many conditional queries to μ, and returns, with probability at least $1-\delta$, a sequence of s samples from a distribution $\bar{\mu}$, where $P=\{\frac{(1+\epsilon)^{k-1}\epsilon}{n}:1\leq k\leq t\}$ for $t=\log n\log(\epsilon^{-1})/\log^2(1+\epsilon)$.*

14.1 Proving the main lemma

The proof of Lemma 42 depends on the following technical lemma.

Lemma 48 *Let μ be a distribution over $[n]$, and let $\mathcal{P}=(P_0,P_1,P_2,\ldots,P_r)$ be a partition of $[n]$ into $r+1$ subsets with the following properties.*

1. *For each $P_k\in\mathcal{P}$, $\mu(i)=p_k$ for every $i\in P_k$, where $p_1,\ldots,p_r$ (but not p_0) are known.*
2. *Given an $i\in[n]$ sampled from μ, we can find the k such that $i\in P_k$.*

Using $s=\frac{6r}{\epsilon^2}\log\left(\frac{r}{\delta}\right)$ samples from μ, with probability at least $1-\delta$, we can find $m_1,\ldots,m_r$ such that $m_k\leq|P_k|$ for all $k\in[r]$, and $\sum_{k\in[r]}p_k(|P_k|-m_k)\leq 4\epsilon$.

Proof Take s samples from μ. For each $k \in [r]$, let s_k be the number of samples in P_k, each with probability p_k. We can easily see that $E[s_k] = sp_k|P_k|$.

If $p_k|P_k| \geq 1/r$, then by Lemma 1, we know that

$$\Pr\left[(1-\epsilon)E[s_k] \leq s_k \leq (1+\epsilon)E[s_k]\right] \geq 1 - 2e^{-\epsilon^2 E[s_k]/3}$$

By the choice of s, with probability at least $1 - \delta/r$, $(1-\epsilon)p_k|P_k| \leq \frac{s_k}{s} \leq (1+\epsilon)p_k|P_k|$.

On the other hand, if $p_k|P_k| < 1/r$, then by Lemma 1 we have

$$\Pr\left[p_k|P_k| - \tfrac{\epsilon}{r} \leq \tfrac{s_k}{s} \leq p_k|P_k| + \tfrac{\epsilon}{r}\right] =$$

$$\Pr\left[\left(1 - \tfrac{\epsilon}{rp_k|P_k|}\right) sp_k|P_k| \leq s_k \leq \left(1 + \tfrac{\epsilon}{rp_k|P_k|}\right) sp_k|P_k|\right] \geq$$

$$1 - 2\exp\left(-\frac{\epsilon^2 s}{3r^2 p_k|P_k|}\right) \geq 1 - 2e^{-\epsilon^2 s/3r}$$

By the choice of s, with probability at least $1 - \delta/r$, $p_k|P_k| - \frac{\epsilon}{r} \leq \frac{s_k}{s} \leq p_k|P_k| + \frac{\epsilon}{r}$.

With probability at least $1 - \delta$, we get an estimate $\alpha_k = s_k/s$ for every $k \in [r]$ satisfying that $\alpha_k \leq \max\{p_k|P_k| + \epsilon/r, p_k|P_k|(1+\epsilon)\}$, and $\alpha_k \geq \min\{p_k|P_k| - \epsilon/r, p_k|P_k|(1-\epsilon)\}$. From now on we assume that α_k satisfies the above bounds, and define $\alpha'_k = \min\{\alpha_k - \epsilon/r, \alpha_k/(1+\epsilon)\}$. Notice that $\alpha'_k \leq p_k|P_k|$. Furthermore, $\alpha'_k \geq \min\{p_k|P_k| - 2\epsilon/r, (1-2\epsilon)p_k|P_k|\}$. Set $m_k = \lceil \alpha'_k/p_k \rceil$. Since $\alpha'_k/p_k \leq |P_k| \in \mathbb{Z}$, we have $m_k \leq |P_k|$ for all k.

Now, $\sum_{k\in[r]} p_k(|P_k| - m_k) \leq \sum_{k\in[r]}(p_k|P| - \alpha'_k)$. Under the above assumption on the value of α_k, for every k such that $p_k|P_k| < 1/r$, we have $\alpha'_k \geq p_k|P_k| - 2\epsilon/r$. Hence, this difference is at most $2\epsilon/r$. For every k such that $p_k|P_k| \geq 1/r$, $\alpha'_k \geq (1-2\epsilon)p_k|P_k|$. For any such k, the difference $p_k|P_k| - \alpha'_k$ is at most $2\epsilon p_k|P_k|$. Therefore, $\sum_{k\in[r]} p_k(|P_k| - m_k)$ is at most 4ϵ.

Proof (Proof of Lemma 42) Given a distribution μ and an interval partition $\mathcal{I} = (I_1, I_2, \ldots, I_r)$, let $\bar{\mu}$ be the distribution presented by the $\epsilon/8$-trimming sampler in Lemma 47. Let $I_{j,k} \subseteq [n]$ be the set of indexes i such that $i \in I_j$ and $\bar{\mu}(i) = \frac{(1+\epsilon/8)^{k-1}\epsilon}{8n}$, and $I_{j,0}$ be the set of indexes in I_j such that $\bar{\mu}(i) = 0$. Thus, each interval I_j in $\mathcal{I}$ is now split into subsets $I_{j,0}, I_{j,1}, I_{j,2}, \ldots, I_{j,\ell}$, where $\ell \leq \log n \log(8/\epsilon)/\log^2(1+\epsilon/8)$.

Using Lemma 47 to obtain $s = r \cdot \mathrm{poly}(\log n, 1/\epsilon, \log(1/\delta))$ samples from the distribution $\bar{\mu}$ (later we calculate how many samples this requires from μ), we can estimate using Lemma 48 where $P_0 = \bigcup_{j=1}^r I_{j,0}$ and $P_{j\ell+k} = I_{j,k}$, with probability at least $1 - \delta/2$, the values $m_{j,k}$ such that $m_{j,k} \leq |I_{j,k}|$ for all $k > 0$, and the following holds.

$$\sum_{j,k:k>0} \tfrac{(1+\epsilon/8)^{k-1}\epsilon}{8n} \left(I_{j,k} - m_{j,k}\right) \leq \epsilon/10.$$

For every j, let M_{I_j} be the inventory provided by $m_{j,1}, \ldots, m_{j,\ell}$ and $m_{j,0} = |I_j| - \sum_{k\in[\ell]} m_{j,k}$. Thus, we have a sequence $\tilde{\mathcal{M}}_{\mathcal{I}} = (M_{I_1}, M_{I_2}, \ldots, M_{I_r})$ of inventories, that is $\epsilon/10$-close in value-distance to the corresponding atlas of $\bar{\mu}$ (where we need to add the interval $\{0\}$ to the partition to cover its entire support). Corresponding to $\tilde{\mathcal{M}}_{\mathcal{I}}$, there is a vector $\tilde{\mu}$ that is $\epsilon/10$-close to $\bar{\mu}$. Using Lemma 46, we have a distribution $\hat{\mu}$ that is $\epsilon/2$-close to $\bar{\mu}$. Since the $[n]$ portion of $\bar{\mu}$ is $\epsilon/2$-close to μ, by the triangle inequality, $\hat{\mu}$ is ϵ-close to μ. Thus $\mathcal{A} = (\mathcal{I}, \mathcal{M}_{\mathcal{I}})$, where $\mathcal{M}_{\mathcal{I}}$ is obtained by multiplying all members of $\tilde{\mathcal{M}}_{\mathcal{I}}$ by the same factor used to produce $\hat{\mu}$ from $\tilde{\mu}$, is an atlas for a distribution that is ϵ-close to μ.

By lemma 47, we need $s \cdot \mathrm{poly}(\log n, \log s, 1/\epsilon, \log(1/\delta))$ conditional samples from μ to get s samples from a correct $\bar{\mu}$ with probability at least $1-\delta/2$. Thus we require $r \cdot \mathrm{poly}(\log n, 1/\epsilon, \log(1/\delta))$ conditional samples from μ to construct the atlas $A_{\mathcal{I}}$.

References

1. Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie, *Testing k-wise and almost k-wise independence*, Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing (New York, NY, USA), STOC '07, ACM, 2007, pp. 496–505.
2. Jayadev Acharya, Clément L. Canonne, and Gautam Kamath, *A chasm between identity and equivalence testing with conditional queries*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA (Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim, eds.), LIPIcs, vol. 40, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2015, pp. 449–466.
3. Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath, *Optimal testing for properties of distributions*, Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada (Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, eds.), 2015, pp. 3591–3599.
4. Tugkan Batu, Lance Fortnow, Eldar Fischer, Ravi Kumar, Ronitt Rubinfeld, and Patrick White, *Testing random variables for independence and identity*, 42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA, 2001, pp. 442–451.
5. Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White, *Testing that distributions are close*, 41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA, IEEE Computer Society, 2000, pp. 259–269.
6. Tugkan Batu, Ravi Kumar, and Ronitt Rubinfeld, *Sublinear algorithms for testing monotone and unimodal distributions*, Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004 (László Babai, ed.), ACM, 2004, pp. 381–390.
7. Clément L. Canonne, *A survey on distribution testing: Your data is big. but is it blue?*, Electronic Colloquium on Computational Complexity (ECCC) **22** (2015), 63.
8. Clément L. Canonne, Ilias Diakonikolas, Themis Gouleakis, and Ronitt Rubinfeld, *Testing shape restrictions of discrete distributions*, 33rd Symposium on Theoretical Aspects of Computer Science, STACS 2016, February 17-20, 2016, Orléans, France, 2016, pp. 25:1–25:14.
9. Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah, *On the power of conditional samples in distribution testing*, SIAM J. Comput. **45** (2016), no. 4, 1261–1296.

10. Clément L. Canonne, Dana Ron, and Rocco A. Servedio, *Testing probability distributions using conditional samples*, SIAM J. Comput. **44** (2015), no. 3, 540–616.
11. Ilias Diakonikolas, *Learning structured distributions*, Handbook of Big Data (2016), 267.
12. Ilias Diakonikolas and Daniel M. Kane, *A new approach for testing properties of discrete distributions*, IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA, 2016, pp. 685–694.
13. Ilias Diakonikolas, Homin K. Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco A. Servedio, and Andrew Wan, *Testing for concise representations*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings, 2007, pp. 549–558.
14. Oded Goldreich and Dana Ron, *On testing expansion in bounded-degree graphs*, Electronic Colloquium on Computational Complexity (ECCC) **7** (2000), no. 20.
15. Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR 0144363
16. Piotr Indyk, Reut Levi, and Ronitt Rubinfeld, *Approximating and testing k-histogram distributions in sub-linear time*, Proceedings of the 31st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2012, Scottsdale, AZ, USA, May 20-24, 2012, 2012, pp. 15–22.
17. Liam Paninski, *A coincidence-based test for uniformity given very sparsely sampled discrete data*, IEEE Transactions on Information Theory **54** (2008), no. 10, 4750–4755.
18. Rocco A. Servedio, *Testing by implicit learning: A brief survey*, Property Testing - Current Research and Surveys [outgrow of a workshop at the Institute for Computer Science (ITCS) at Tsinghua University, January 2010], 2010, pp. 197–210.
19. Gregory Valiant and Paul Valiant, *A CLT and tight lower bounds for estimating entropy*, Electronic Colloquium on Computational Complexity (ECCC) **17** (2010), 179.
20. ———, *An automatic inequality prover and instance optimal identity testing*, SIAM J. Comput. **46** (2017), no. 1, 429–455.