Testing vs Estimation for Index-Invariant Properties in the Huge Object Model

Sourav Chakraborty[*]   Eldar Fischer[†]   Arijit Ghosh[‡]   Amit Levi[§]

Gopinath Mishra[¶]   Sayantan Sen[‖]

Abstract

The Huge Object model of property testing [Goldreich and Ron, TheoretiCS 23] concerns properties of distributions supported on $\{0,1\}^n$, where n is so large that even reading a single sampled string is unrealistic. Instead, query access is provided to the samples, and the efficiency of the algorithm is measured by the total number of queries that were made to them.

Index-invariant properties under this model were defined in [Chakraborty et al., COLT 23], as a compromise between enduring the full intricacies of string testing when considering unconstrained properties, and giving up completely on the string structure when considering label-invariant properties. Index-invariant properties are those that are invariant through a consistent reordering of the bits of the involved strings.

Here we provide an adaptation of Szemerédi's regularity method for this setting, and in particular show that if an index-invariant property admits an ϵ-test with a number of queries depending only on the proximity parameter ϵ, then it also admits a distance estimation algorithm whose number of queries depends only on the approximation parameter.

---

[*]Indian Statistical Institute, Kolkata, India. Research partially supported by the SERB project SQUID-1982-SC-4837. Email: chakraborty.sourav@gmail.com.

[†]Technion - Israel Institute of Technology, Israel. Email: eldar@cs.technion.ac.il. Research supported by an Israel Science Foundation grant number 879/22.

[‡]Indian Statistical Institute, Kolkata, India. Research partially supported by the Science & Engineering Research Board of the DST, India, through the MATRICS grant MTR/2023/001527. Email: arijitiitkgpster@gmail.com.

[§]University of Haifa, Israel. Email: alevi@cs.haifa.ac.il.

[¶]National University of Singapore, Singapore. Email: gopinath@nus.edu.sg.

[‖]Centre for Quantum Technologies, National University of Singapore, Singapore. Research supported by the National Research Foundation, Singapore and A*STAR under its Quantum Engineering Programme NRF2021-QEP2-02-P05. The author would like to thank Eldar Fischer for hosting him at Technion for an academic visit where this work was initiated. Email: sayantan789@gmail.com.

Contents

1 Introduction

Distribution testing is a sub-field of property testing [RS96, GGR98] which has received a significant amount of attention in the past few decades (See e.g., [Fis04, Ron08, Ron10, CS10, RS11, Gol17, Can20, Can22, BY22]). In the classical setting of distribution testing, an algorithm can draw independent samples from an unknown input distribution μ, and needs to either accept or reject it based on the drawn samples. An ϵ-testing algorithm for some distribution property $\mathcal{P}$ is required to accept every distribution satisfying $\mathcal{P}$ with high probability, and reject every distribution that has distance of at least ϵ from any distribution satisfying $\mathcal{P}$ with high probability, while drawing as few samples as possible.

One particular setting of interest, mainly due to its resemblance to real-world data, is the setting where the distribution is supported on a high dimensional product space. In the classical setting of distribution testing, it is implicitly assumed that the algorithm has complete access to the sampled elements. However, for product spaces with extremely large dimension, such access is impractical: Even obtaining the entire label of a single drawn element might be too expensive.

To handle this setting, Goldreich and Ron [GR23] defined and initiated the *Huge Object model*. In this model the algorithm can draw samples from the distribution, but since the samples are not fully accessible, the algorithm is provided with query access to the sampled strings. The distance notion is changed accordingly from the total variation distance to the *Earth Mover's distance*. Various properties and algorithmic behaviors have been studied under the huge object model (See [GR23, CFG+23, AF24, AFL24, CSY25]).

When considering properties of distributions, it is commonly the case that the property at hand has some "symmetry" in its structure. One such symmetry is *label-invariance*. A property is label-invariant if applying any permutation to the labels does not effect membership in the property. Label-invariant properties have been extensively studied in the literature [BDKR05, Pan08, GR11, Val11, DKN14, CDVV14, ADK15, VV17, BC17, DKS18, CFG+22].

Recently, Chakraborty et al. [CFG+23] considered the general task of testing *index-invariant* properties under the huge object model. A property of distributions supported on $\{0,1\}^n$ is called index-invariant if any permutation $\pi : [n] \to [n]$ of the indices of the strings does not effect membership in the property. Note that the class of index-invariant properties extends the class of label-invariant properties, since the invariance condition is weaker compared to label-invariance: It only considers relabeling strings by rearranging their symbols through a common permutation of the indices. In particular, index-invariance still allows for properties to have some string-related structure. For example, monotonicity of the probability mass function with respect to the natural partial order on $\{0,1\}^n$ is an index-invariant property but not a label-invariant one.

The main result of [CFG+23] showed that the number of queries is controlled through upper and lower bounds depending on the VC-dimension of the support of the distributions in the property (along with the proximity parameter). In addition, they showed that for properties that are not index-invariant such bounds are not guaranteed, and that there exists an almost tight quadratic gap between adaptive testers (i.e., the setting where queries might depend on answers to previous queries) and non-adaptive testers (i.e., the setting where queries are determined in advanced) of index-invariant properties. Note that such a distinction is non-existent in the classical model of distribution testing, as there is no notion of adaptivity when the algorithm is only allowed to draw independent samples from the distribution.

In this work, we complement the result of [CFG+23] and show that for any index-invariant property that is ϵ-testable using a constant (depending only on ϵ) number of queries, there exists an estimation algorithm that approximates the distance of the input distribution from the property, by performing only a constant (depending only on ϵ) number of queries.

Theorem 1 (Informal statement of Theorem 2). *If $\mathcal{P}$ is an index-invariant property of distributions over $\{0,1\}^n$ which is ϵ'-testable with a constant number of queries for any fixed $\epsilon' > 0$, then there is an estimation algorithm with a constant number of queries for any fixed $\epsilon > 0$, that for a distribution μ of distance d from $\mathcal{P}$, outputs a distance approximation $\boldsymbol{\eta}$ such that with high probability $|\boldsymbol{\eta} - d| \leq \epsilon$.*

This brings index-invariant properties under the umbrella of properties for which constant-query testability implies constant-query estimability, similar to dense graph properties, and in contrast to general string properties and also general Huge Object model properties. Along the way, we provide a "regularity-like" framework to index-invariant properties, to treat them using a "meta-testing" scheme. More on this will be discussed in Section 1.1. It is important to note that the dependencies here are not a tower function of $1/\epsilon$, a distinction from most regularity-like frameworks in the literature, which usually provide a dependency which is a tower in the other parameters, or even worse. In particular, the query complexity in our result has double-exponential dependency on $1/\epsilon$. Section 1.2 contains more information on this and relations to prior works.

1.1 Technical overview

The proof of our main result is motivated by Szemerédi's regularity lemma for graphs [Sze75]. Intuitively, Szemerédi's regularity lemma states that a graph can be partitioned into a bounded number of parts such that the subgraphs between the parts are "random-like". In addition, we take inspiration from [FN07b] that proves an analogous estimation result for the dense graph testing model, and in particular we define and use a notion inspired by the notion of robust partitions from [FN07b].

1.1.1 Detailings

Our starting point is to define a notion that resembles that of a graph partition for the setting of distributions. For technical reasons, we define a notion that is slightly more generalized. A *detailing* ξ of a distribution μ over $\{0,1\}^n$ with respect to some set A is a distribution on $\{0,1\}^n \times A$ such that the distribution on $\{0,1\}^n$ that is projected from ξ is equal to μ. Loosely speaking, a detailing of μ with respect to A is in particular a representation of μ as a weighted sum of distributions, where the weights are given by the distribution over A projected from ξ, and for every $a \in A$ the corresponding "component distribution" μ_a results from conditioning ξ to the event $\{0,1\}^n \times \{a\}$.

One particular type of a detailing which we consider extensively in this work is a *detailing by a set of variables*. For a set $U \subseteq \{1, \ldots, n\}$ the detailing of μ with respect to U is a distribution μ^U supported over $\{0,1\} \times A$ for $A = \{0,1\}^U$, where for $a \in A$ the component μ_a is exactly the conditioning of μ to the event $\{x \in \{0,1\}^n : x_U = a\}$. Note that this detailing is actually a partition of the probability space to weighted subcubes indexed by assignments to variables in U. A detailing by variables can be better handled by a testing algorithm as compared to a general detailing, because a draw $(\boldsymbol{x}, \boldsymbol{a}) \sim \mu^U$ can be easily obtained by taking a sample $\boldsymbol{x} \in \{0,1\}^n$ from

μ and augmenting it with $\boldsymbol{a} = \boldsymbol{x}_U$. In comparison, a draw from a general detailing ξ is not always easily obtainable given only samples from the input distribution μ.

The notion of a detailing allows us to extract some useful *statistics* of the underlying distribution. The statistic of a detailing ξ with respect to A is essentially the distribution η over A that is the projection of ξ to the second argument, along with a multiset of *type* vectors $\{t_1, \ldots, t_n\}$ over $[0, 1]^A$ where $t_i(a) = \mathbf{Pr}_{\boldsymbol{x} \sim \mu_a}[\boldsymbol{x}_i = 1]$. Namely, $t_i(a)$ is the probability that the variable $\boldsymbol{x}_i$ is set to one conditioned on the event that a draw from ξ lies in $\{0, 1\}^n \times \{a\}$. Equivalently, the multiset of type vectors can be identified with a distribution Λ supported on vectors in $[0, 1]^A$. We refer to η as the *weight distribution*, and to Λ as the *type distribution* of the detailing ξ. Intuitively, one can consider the statistic as a sort of a "lossy compression scheme" for the underlying distribution.

1.1.2   Goodness and predictability

One important property that a detailing can have is being (ϵ, q)-*good*. For $\epsilon \in (0, 1)$ and $q \in \mathbb{N}$, we say that a single distribution ν is (ϵ, q)-good if at least a $1 - \epsilon$ fraction of the q-tuples of variables under this distribution are ϵ-close to being independent (this is a weak form of being "somewhat q-wise independent"). We say that a detailing ξ of μ is (ϵ, q)-good if at least a $1 - \epsilon$ fraction of the component distributions μ_a (counted by their weights in the detailing) are (ϵ, q)-good.

This property is extremely beneficial for finding a distribution that approximates the distribution of samples and queries made by a *canonical algorithm* using the statistic of ξ. A canonical algorithm for a distribution property $\mathcal{P}$ is a randomized procedure that picks a set of s independent samples drawn from the distribution μ, and a set of q variables (that is "coordinates" from the space $\{0, 1\}^n$, identified by their indices from $\{1, \ldots, n\}$), chosen uniformly at random, and makes its decision (either accept or reject) based on the samples restricted to the chosen variables (and possibly on its internal random coins). In particular, the acceptance probability of the tester is determined by a function $\alpha : \{0, 1\}^{s \times q} \to [0, 1]$. Considering canonical testers suffices for our needs, since for any index-invariant property having a testing algorithm making s samples and q queries, there exists a corresponding canonical tester which makes at most sq queries, by [CFG+23].

The goodness property allows us to simulate the behaviour of a canonical algorithm by using only the detailing statistic (this is essentially the "meta-testing" concept that was featured in [FN07b, AFNS06]). In particular, to simulate the behaviour of a canonical algorithm with s samples and q queries we first draw $\boldsymbol{a}_1 \ldots, \boldsymbol{a}_s$ independently from the weight distribution η of the detailing, and then draw q vectors $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_q$ from Λ independently. Then, we draw a random matrix $\mathbf{M} \in \{0, 1\}^{s \times q}$ such that for every $(i, j) \in \{1, \ldots, s\} \times \{1, \ldots, q\}$, the value $\mathbf{M}_{i,j}$ is drawn independently according to the Bernoulli distribution with parameter $\boldsymbol{t}_j(\boldsymbol{a}_i)$.

Essentially, for every $i \in \{1, \ldots, s\}$ the draw $\boldsymbol{a}_i$ simulates the choice of the respective component distribution $\mu_{\boldsymbol{a}_i}$ of ξ, and the values $\boldsymbol{t}_1(\boldsymbol{a}_i), \ldots, \boldsymbol{t}_q(\boldsymbol{a}_i)$ simulate the queries to the sample i. Note that since $\boldsymbol{t}_j(\boldsymbol{a}_i)$ corresponds to the j-th marginal of a draw from the component distribution $\mu_{\boldsymbol{a}_i}$, and the drawn component distribution with high probability satisfies the goodness condition, the distribution of the string $\mathbf{M}_{i,1}, \ldots, \mathbf{M}_{i,q}$ is close to the distribution of a sample drawn from μ and restricted to the set of q uniformly random indices. The matrix $\mathbf{M}$ is then used along with α to determine the acceptance probability of the algorithm. This implies a *predictability* property of the detailing: One can approximate the acceptance probability of a canonical algorithm by using only its statistic and without actually performing the queries to the input.

1.1.3 Robustness and inheritance

The *predictability* notion is tied to a more general notion of *robustness*. Roughly speaking, a detailing is robust if whenever it is extended to a more refined detailing (but still with a limited degree of refinement), the perceived probabilities of the outcomes in individual indices will not change by much. In other words, for the most parts, the values appearing in the type distribution of the extended detailing will be *inherited* from their counterparts in the type distribution of the original detailing. Such an inheritance feature will be crucial for designing the distance estimation algorithm. To facilitate such a feature, instead of tracking all the individual probabilities we go the route of Szemerédi's regularity lemma, and define a single "detailing-index" measure for a detailing, with respect to which robustness is defined. The above robustness notion is then proved to follow from satisfying the detailing-index-related measure.

However, due to the access we have to the input we are unable to find such a robust detailing. We can only consider extensions which are obtained by fixing a subset of the bits in the string. As a result, we define a weaker notion of robustness. A detailing ξ is said to be *weakly robust* if extending it to a more refined detailing obtained by fixing a randomly chosen and not too large subset of the bits (as opposed to arbitrary refinements in the general definition of robustness) will not change the perceived probabilities of the outcomes in individual indices by much. We later prove that being a weakly robust detailing implies the general notion of robustness, and implies the above mentioned goodness notion as well.

1.1.4 Distance estimation algorithm

A main engine of our work is an efficient query algorithm that provides an approximation of the statistic of a robust detailing. We set the robustness requirement to go beyond what is needed just for predictability, as the above-mentioned inheritance feature will be required for distance estimation later on. Given the detailing statistic, we then consider hypothetical distributions with hypothetical detailings featuring predictability.

For such a hypothetical detailing we first require that its statistic must predict acceptance (or at least predict a high enough probability of not rejecting) by a testing algorithm. For a statistic satisfying this requirement, we check its hypothetical distance from the distribution (which we actually measure using a distance measure between the corresponding type statistics). The smallest such hypothetical distance gives us our estimate.

On the one hand, if there exists a relatively close distribution that satisfies the property, it would imply a corresponding hypothetical detailing statistic: We consider a predictive detailing of this target distribution, and show how to "combine" it with the detailing we obtained from the input to demonstrate a distance close to the hypothetical one. Here the inheritance property is crucial, since the combined detailing is related to a refinement of the detailing of the input, and we need it to maintain its type statistic.

On the other hand, if there exists a good hypothetical detailing statistic with a small hypothetical distance, it would imply the existence of a "fake" distribution that has a similar distance bound from the input distribution, and additionally is not rejected by the test (and hence there must be a "real" distribution satisfying the property that is not much further away). Building such a distribution is mostly a matter of considering a draw from a detailing, and then altering each coordinate independently with some probability that is tailored to eventually match the hypothetical

acceptable statistic.

1.2 Relation to prior works

The study of tolerant property testing and distance approximation (also known as estimation) was initiated by Parnas, Ron and Rubinfeld [PRR06]. In the setting of tolerant testing, the algorithm is required to accept with high probability if the input is close to the property, and reject with high probability if the input is far from any other input satisfying the property. The relation between "non-tolerant" property testing and distance approximation depends much on whether invariance restrictions are imposed on a property. While testable graph properties (which must be invariant with respect to graph isomorphism) were proved to admit full distance estimation in [FN07b], in the general string testing model (where the properties are not required to satisfy any invariance condition) the existence of non-tolerantly testable properties which are not estimable, and in fact not even tolerantly testable, was proved in [FF06]. Later, it was proved in [BFLR20] that there are testable properties that require a near-linear number of queries for a tolerant test.

The results here exhibit the role that invariance requirements can play in the setting of the Huge Object Model. In the most general Huge Object Model setting, features of traditional string testing factor in, and in particular one can use the reductions from the original [GR23] to convert the properties from [FF06] or [BFLR20] to create non-tolerantly testable properties which are not tolerantly testable in the Huge Object model (as mandated by Theorem 1, these properties are indeed not index-invariant).

In [GR23] two invariance notions were considered. These are label-invariance which discards most of the "string-ness" of this model and focuses on distribution properties (such as uniformity and bounded support size) that just happen to be represented by distributions on strings, and the even more restrictive notion of mapping-invariance. In [CFG+23] a study of a milder notion of invariance was initiated, namely the notion of *index-invariance*. This work proves that index-invariant properties already belong to the domain where testability implies estimability, as is the case with the dense graph model.

The influence of Szemerédi's regularity-like constructions in property testing also has a long history, starting with [AFKS00] (in fact a mathematical result in this direction has already been presented in [RD85]). Such constructions were usually required to go beyond the original regularity lemma, at times in terms of strength (as in [AFKS00]) and at other times in terms of the objects involved outside standard graphs, such as product posets [FN07a], graphs in the sparse model [BS11, BCCZ19, BCCZ18, CFSZ21] or vertex-ordered graphs [ABEF17].

To avoid ad-hoc constructions, there are several mathematical approaches to systematically strengthening the notion of regularity and defining it for other objects, and two of them gained particular prominence. The most influential one is the analytic approach, developed in [LS06, BCL+06a, BCL+06b, LS07, BCL+08, LS10, BCL+12], which has many applications to other research domains (see e.g., [HKM+13, CR20, GHHS20, Sim23]). For an extensive introduction to the analytic approach see [Lov12]. This approach works particularly well in settings where the exact dependency functions do not matter. The other approach is that of robust partitions (or objects) which was first used in [FN07b]. Many times both approaches work, compare e.g. the robustness approach of [ABEF17] with its analytic counterpart [BEFLY21, GHHS20]. It is the robustness approach that we use here.

A disconcerting feature of regularity-related approaches is that they almost always provide parameter dependencies that are a tower function or even worse (e.g. tower of towers). The analytic approach tends to do away with explicit dependencies altogether. However, there are exceptions, and much research went into reducing the dependencies in cases that allow it, starting with [FK99]. Of particular interest is the work of [GKS23], which provided a version of the estimability result from [FN07b] whose parameter dependency is not a tower. Also in this work the resulting dependency is not a tower, although its (constant) number of exponentiations leaves something to be desired. Considering the lower bounds in [CFG+23] related to the VC dimension, it is unlikely that this dependency can be reduced.

Another interesting insight from the comparison with the work of [GKS23] is gained by noting that in our work the regularity framework in itself has a non-tower dependency, while for the dense graph setting this is not possible by the lower bound of [Gow97] (see also [MS16]). In [GKS23] they manage to move the specific estimability proof to use the weak regularity framework of [FK99] instead.

1.3   Organization of the paper

We start with somewhat longer than usual preliminaries, Section 2, that provide the basic groundwork for our handling of distributions. Then, in Section 3 we develop our definitions of a detailing of a distribution, and the definition for its goodness and robustness. In Section 4 we show how from only knowing the statistic of a good detailing we can predict the behavior of a canonical test over a given input distribution, while in Section 5 we show how to find a robust (and good) detailing by variables, and estimate its statistic. The final Section 6 is where all of this comes together in our distance estimation algorithm.

2   Preliminaries

2.1   Basic handling and manipulation of distributions

For an integer n, we will denote the set $\{1, \ldots, n\}$ as $[n]$. We use boldface letters (such as $\boldsymbol{x}$) to denote random variables. Given two vectors x and y in $\{0, 1\}^n$, we denote by $d_H(x, y)$ the normalized Hamming distance between x and y, that is, $d_H(x, y) \stackrel{\text{def}}{=} \frac{1}{n} \cdot |\{i \in [n] : x_i \neq y_i\}|$.

Given a discrete distribution μ over Ω, for $x \in \Omega$ we use $\mathbf{Pr}_\mu[x]$ and $\mu(x)$ interchangeably (while for an event $S \subseteq \Omega$ we only use $\mathbf{Pr}_\mu[S] = \sum_{x \in S} \mu(x)$ for its probability). For $p \in [0, 1]$, we let $\mathrm{Ber}(p)$ denote the Bernoulli distribution with parameter p.

To streamline the arguments and analysis in the following, we will define and use a somewhat uncommon notation for some very common probabilistic notions.

Definition 2.1 (Restriction of a distribution). Let μ be a distribution over Ω and $S \subseteq \Omega$ be an event such that $\mathbf{Pr}_\mu[S] \neq 0$. Then $\mu|^S$ is the distribution μ restricted to S, which is defined over S as follows: For every $x \in S$, $\mathbf{Pr}_{\mu|^S}[x] = \frac{\mathbf{Pr}_\mu[x]}{\mathbf{Pr}_\mu[S]}$. We also sometimes pad the conditional distribution to Ω by defining $\mu|^S(x) = 0$ for $x \in \Omega \setminus S$.

Definition 2.2 (Projection of a distribution). Let μ be a distribution over $\Omega = \prod_{\ell=1}^n A_{i_\ell}$ and $i_1, \ldots, i_k$ be a set of integers from $[n]$ such that $i_1 < \cdots < i_k$. The *projection* $\mu|_{\{i_1, \ldots, i_k\}}$ of μ

to a set of coordinates $\{i_1, \ldots, i_k\}$ is the distribution supported on $\prod_{\ell \in [k]} A_{i_\ell}$ obtained by first drawing $\boldsymbol{x} \sim \mu$ and returning $\boldsymbol{x}|_{\{i_1, \ldots, i_k\}}$. Moreover, for an event $S \subseteq \Omega$, we denote by $\mu|_{\{i_1, \ldots, i_k\}}^S$ the conditioning of μ over S followed by the projection on $\{i_1, \ldots, i_k\}$.

Definition 2.3 (Conditioning shorthand). When we have a distribution μ on a set Ω which is a product, we use a shorthand notation when we condition it by the coordinate values. For example, for $\Omega = A \times B$ and $a \in A$, we use $\mu|^{1:a}$ to denote the conditioning $\mu|^{\{a\} \times B}$, and similarly for $A' \subseteq A$ use $\mu|^{1:A'}$ to denote $\mu|^{A' \times B}$. We extend this notation to multiple coordinates as well. For example, for $\Omega = A \times B \times C$ and $a \in A$ and $c \in C$, we use $\mu|^{1,3:(a,c)}$ or $\mu|^{1:a,3:c}$ to denote $\mu|^{\{a\} \times B \times \{c\}}$.

For two distributions that "agree on a common variable" we define a way to unify them into a single distribution, named after the database operation resembling it.

Definition 2.4 (Join of distributions). Given two distributions μ over $A \times B$ and ν over $B \times C$ that satisfy $\mu|_2 = \nu|_1$, we define their *join* $\mu \bowtie \nu$, as the following distribution over $A \times B \times C$:

$$(\mu \bowtie \nu)(a, b, c) = \mu(a, b) \cdot \nu|_2^{1:b}(c) = \mu|_1^{2:b}(a) \cdot \nu(b, c).$$

Whenever $\mu|_2(b) = \nu|_1(b) = 0$ we (as expected) define $(\mu \bowtie \nu)(a, b, c) = 0$. In cases where it is not clear which coordinate we unify (for example when $A = B$) we state it explicitly.

We also define a way to "adjust" a distribution over a product set to have a desired projections to one of its coordinates.

Definition 2.5 (Adjustment). Given a distribution η over $A \times B$ and a distribution ν over A, such that $\nu(a) = 0$ whenever $\eta|_1(a) = 0$, the *adjustment of η to ν*, denoted $\nu \triangleright \eta$, is defined as the following distribution over $A \times B$:

$$(\nu \triangleright \eta)(a, b) = \nu(a) \cdot \eta|_2^{1:a}(b).$$

Whenever $\eta|_1(a) = 0$ (and then also $\nu(a) = 0$) we as expected define $(\nu \triangleright \eta)(a, b) = 0$.

2.2 Distribution distances

The following is the most basic distance notion between distributions.

Definition 2.6 (Variation distance). Let μ and τ be two probability distributions over a discrete set Ω. The *variation distance* between μ and τ is defined as:

$$d_{\mathrm{TV}}(\mu, \tau) \stackrel{\text{def}}{=} \frac{1}{2} \sum_{w \in \Omega} |\mu(w) - \tau(w)|.$$

Central to our work will be the notion of Earth Mover distance, with some variants. Before defining it, let us consider the notion of transfer distribution.

Definition 2.7 (Transfer distribution). Let μ and τ be two distributions defined over two sets A and B, respectively. A distribution T over $A \times B$ is said to be a *transfer distribution* between μ and τ if for every $a \in A$, $\mathbf{Pr}_{(\boldsymbol{x}, \boldsymbol{y}) \sim T}[\boldsymbol{x} = a] = \mu(a)$, and for every $b \in B$, $\mathbf{Pr}_{(\boldsymbol{x}, \boldsymbol{y}) \sim T}[\boldsymbol{y} = b] = \tau(b)$ (in other words, $T|_1 = \mu$ and $T|_2 = \tau$). The set of all transfer distributions between μ and τ is denoted by $\mathcal{T}(\mu, \tau)$.

Definition 2.8 (EMD with respect to a distance function). Let μ and τ be two distributions defined over a set Ω and d_Ω be a distance function defined over Ω. Then the *Earth Mover distance* (EMD) between μ and τ with respect to d_Ω is defined as follows:

$$d_{\mathrm{EM}}(\mu, \tau) \overset{\mathrm{def}}{=} \inf_{T \in \mathcal{T}(\mu,\tau)} \operatorname*{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y}) \sim T} [d_\Omega(\boldsymbol{x}, \boldsymbol{y})]$$

where $\mathcal{T}(\mu, \tau)$ denotes the set of all possible transfer distributions between μ and τ.

Unless stated otherwise, whenever $\Omega = \{0, 1\}^n$, we use d_{EM} to refer to the EMD over the normalized Hamming distance.

We will not really have to take care of limits in our arguments due to the following observation.

Observation 2.9. *If μ and τ are distributions over a finite set, then $\mathcal{T}$ is compact and in particular there exists a transfer distribution $T \in \mathcal{T}$ achieving the respective EMD distance.*

The following observation, that variation distance is essentially also a special case of the Earth Mover distance, is well-known.

Observation 2.10. *The Earth Mover distance over the Kronecker distance function is identical to the total variation distance. That is,*

$$d_{\mathrm{TV}}(\mu, \tau) \overset{\mathrm{def}}{=} \inf_{T \in \mathcal{T}(\mu,\tau)} \operatorname*{\mathbf{E}}_{(\boldsymbol{x},\boldsymbol{y}) \sim T} \left[\mathbf{1}_{\{\boldsymbol{x} \neq \boldsymbol{y}\}} \right],$$

where $\mathbf{1}_{\{\boldsymbol{x} \neq \boldsymbol{y}\}}$ is the indicator function for the event that $\boldsymbol{x} \neq \boldsymbol{y}$.

Keeping with tradition, in the variation distance setting we will also refer to a transfer distribution as a *coupling*. The following lemma (which will be useful to us in the sequel) is a good example of an application of the above observation.

Lemma 2.11. *If ν is a distribution over A and η is a distribution over $A \times B$, then the adjustment of η to ν (see Definition 2.5) satisfies $d_{\mathrm{TV}}(\eta, \nu \triangleright \eta) \leq d_{\mathrm{TV}}(\eta|_1, \nu)$.*

Proof: Assume that T is an optimal coupling between $\eta|_1$ and ν. That is, $\operatorname*{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{a}') \sim T} \left[\mathbf{1}_{\{\boldsymbol{a} \neq \boldsymbol{a}'\}} \right] = d_{\mathrm{TV}}(\eta, \nu)$. We use it to define a (not necessarily optimal) coupling T' between η and $\nu \triangleright \eta$ for which $d_{\mathrm{TV}}(\eta, \nu \triangleright \eta) \leq \operatorname*{\mathbf{E}}_{((\boldsymbol{a},\boldsymbol{b}),(\boldsymbol{a}',\boldsymbol{b}')) \sim T'} \left[\mathbf{1}_{\{(\boldsymbol{a},\boldsymbol{b}) \neq (\boldsymbol{a}',\boldsymbol{b}')\}} \right] \leq \operatorname*{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{a}') \sim T} \left[\mathbf{1}_{\{\boldsymbol{a} \neq \boldsymbol{a}'\}} \right]$, which completes the proof.

A draw $((\boldsymbol{a},\boldsymbol{b}),(\boldsymbol{a}',\boldsymbol{b}')) \sim T'$ is taken as follows: We first draw $(\boldsymbol{a},\boldsymbol{a}') \sim T$. If $\boldsymbol{a} = \boldsymbol{a}'$, then we let $\boldsymbol{b} = \boldsymbol{b}'$ be the result of a single draw from $\eta|_2^{1:\boldsymbol{a}}$. If $\boldsymbol{a} \neq \boldsymbol{a}'$, then we let $\boldsymbol{b}$ be the result of a draw from $\eta|_2^{1:\boldsymbol{a}}$, and let $\boldsymbol{b}'$ be the result of an independent draw from $\eta|_2^{1:\boldsymbol{a}'}$.

It is not hard to see that T' is indeed a transfer distribution between between η and $\nu \triangleright \eta$. For example, $T'|_1(a, b) = T|_1(a) \cdot \eta|_2^{1:a}(b) = \eta|_1(a) \cdot \eta|_2^{1:a}(b) = \eta(a, b)$, and the proof for $T'|_2 = \nu \triangleright \eta$ is similar. It remains to bound the probability for the event $(\boldsymbol{a}, \boldsymbol{b}) \neq (\boldsymbol{a}', \boldsymbol{b}')$, and this follows from noting that the definition of T' explicitly states that $\boldsymbol{a} = \boldsymbol{a}'$ implies $\boldsymbol{b} = \boldsymbol{b}'$. ∎

In the sequel, we will also use the Earth Mover distance over the weighted ℓ_1 norm.

Definition 2.12 (EMD with respect to weighted ℓ_1-distance). Let η be a distribution over A. Also, let x and y be two vectors in $[0,1]^A$. The η-weighted ℓ_1-distance between x and y is defined as

$$d_{\ell_1}^{\eta}(x,y) \stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \eta}[|x_{\boldsymbol{a}} - y_{\boldsymbol{a}}|].$$

Considering two distributions Λ and Υ defined over $[0,1]^A$, the EMD between them with respect to the η-weighted ℓ_1-distance is denoted by $d_{\text{EM}}^{\eta}(\Lambda, \Upsilon)$.

In our setting, the distributions Λ and Υ will always be finitely-supported (even that Ω itself is infinite), and in particular, Observation 2.9 will still apply.

2.3   The testing model

The Huge Object model uses the following oracle access to the unknown input distribution.

Definition 2.13 (Huge Object oracle). In the Huge Object model, the algorithm can access the input distribution μ in the following manner: At every stage, the algorithm may ask for a new sample $\boldsymbol{x} \sim \mu$, independently of all previous samples, or it may ask to query a coordinate $j \in \{1, \ldots, n\}$ of a previously obtained sample $\boldsymbol{x}'$. When such a query is made, the output of the oracle is $\boldsymbol{x}'_j \in \{0,1\}$.

Definition 2.14 (Distribution Property). A *distribution property* $\mathcal{P}$ is a sequence $\mathcal{P}_1, \mathcal{P}_2, \ldots$ such that for every $n \geq 1$, $\mathcal{P}_n$ is a compact subset of the set of all distributions over $\{0,1\}^n$.

Definition 2.15 (Distance of a distribution from a property). Let $\mathcal{P} = (\mathcal{P}_1, \mathcal{P}_2, \ldots)$ be a property and μ be a distribution over $\{0,1\}^n$ for some n. The *distance of μ from $\mathcal{P}$* is defined as $d_{\text{EM}}(\mu, \mathcal{P}) = \min_{\tau \in \mathcal{P}_n}\{d_{\text{EM}}(\mu, \tau)\}$.

Testing and tolerant testing are defined as follows.

Definition 2.16 (ϵ-tester). Fix $0 < \epsilon \leq 1$ and let $\mathcal{P}$ be a property of distributions supported on $\{0,1\}^n$. An ϵ-*tester* for $\mathcal{P}$ is a randomized procedure that has Huge Object oracle access to an input distribution μ and satisfies the following with probability at least $2/3$:

1. (Completeness) If $\mu \in \mathcal{P}$, then the algorithm outputs **Accept**.

2. (Soundness) If $d_{\text{EM}}(\mu, \mathcal{P}) > \epsilon$, then the algorithm outputs **Reject**.

Definition 2.17 ((ϵ_1, ϵ_2)-tolerant tester). Fix $0 \leq \epsilon_1 < \epsilon_2 \leq 1$ and let $\mathcal{P}$ be a property of distributions supported on $\{0,1\}^n$. An (ϵ_1, ϵ_2)-*tolerant tester* for $\mathcal{P}$ is a randomized procedure that has Huge Object oracle access to an input distribution μ and satisfies the following with probability at least $2/3$:

1. (Completeness) If $d_{\text{EM}}(\mu, \mathcal{P}) \leq \epsilon_1$, then the algorithm outputs **Accept**.

2. (Soundness) If $d_{\text{EM}}(\mu, \mathcal{P}) > \epsilon_2$, then the algorithm outputs **Reject**.

Note that an (non-tolerant) ϵ-tester for a property $\mathcal{P}$ is the same as a "$(0, \epsilon)$-tolerant" tester for $\mathcal{P}$.

In this work we will be interested in the following class of properties.

Definition 2.18 (Index-invariant property). Let μ be a distribution over $\{0,1\}^n$. For any permutation $\pi : [n] \to [n]$, let μ_π be the distribution such that $\mu(x_1, \ldots, x_n) = \mu_\pi(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for every $x \in \{0,1\}^n$. A distribution property $\mathcal{P}$ is called *index-invariant* if $\{\mu_\pi : \mu \in \mathcal{P}\} = \mathcal{P}$ for every permutation π.

2.4 Canonical testers

The following is the definition of the information obtained by a specific sampling and querying pattern from an input distribution μ.

Definition 2.19 (Canonical distribution). Fix $s, q \in \mathbb{N}$ and let μ be a distribution over $\{0,1\}^n$. The (s,q)-*canonical distribution* for μ is a distribution $\mathcal{D}_{\text{test}}^{s,q}$ over $\{0,1\}^{s \times q}$ obtained by the following process. Draw s independent samples $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^s)$ from μ, then pick a uniformly random q-tuple $(\boldsymbol{j}_1, \ldots, \boldsymbol{j}_q) \in [n]^q$, and finally return the following matrix:

$$
\mathbf{M} = \begin{bmatrix} \boldsymbol{x}_{\boldsymbol{j}_1}^1 & \cdots & \boldsymbol{x}_{\boldsymbol{j}_q}^1 \\ \vdots & & \vdots \\ \boldsymbol{x}_{\boldsymbol{j}_1}^s & \cdots & \boldsymbol{x}_{\boldsymbol{j}_q}^s \end{bmatrix} .
$$

When s and q are clear from the context, we omit the superscript and use $\mathcal{D}_{\text{test}}$ to denote the above.

Canonical testers are defined as those testers that sample and query according to the above pattern.

Definition 2.20 ((s,q)-canonical tester). Fix $\epsilon \in (0,1)$. An (s,q)-canonical tester with proximity parameter ϵ for some distribution property $\mathcal{P}$, is a randomized procedure that acts by obtaining a matrix $\mathbf{M}$ from the (s,q)-canonical distribution $\mathcal{D}_{\text{test}}$ and then accepting or rejecting based on $\mathbf{M}$ and possibly some internal coin tosses. We let $\alpha_\epsilon : \{0,1\}^{s \times q} \to [0,1]$ so that $\alpha_\epsilon(\mathbf{M})$ denotes the probability that the (s,q)-canonical tester (with proximity parameter ϵ) accepts $\mathbf{M}$. The acceptance probability of the tester is denoted by $\mathsf{acc}_\epsilon(\mu) \stackrel{\text{def}}{=} \mathbf{E}_{\mathbf{M} \sim \mathcal{D}_{\text{test}}}[\alpha_\epsilon(\mathbf{M})]$.

In particular, an (s,q)-canonical ϵ-test must satisfy the following. If $\mu \in \mathcal{P}$ then the tester accepts with probability at least $2/3$ (i.e., $\mathsf{acc}_\epsilon(\mu) \geq 2/3$) and if $d(\mu, \mathcal{P}) > \epsilon$, then the tester rejects with probability at least $2/3$ (i.e., $\mathsf{acc}_\epsilon(\mu) < 1/3$). The total number of queries of such a test is sq.

The following lemma states that for index-invariant properties, general (even adaptive) testers can be converted to canonical testers at a cost that is at most quadratic in their number of queries.

Lemma 2.21 (Theorem 1.7 in [CFG$^+$23]). *Fix $\epsilon \in (0,1)$. Let $\mathcal{P}$ be an index-invariant property such that there exists a (possibly adaptive) ϵ-test for $\mathcal{P}$ with s samples and q queries. Then there exists an (s,q)-canonical ϵ-test for $\mathcal{P}$ performing at most $sq \leq q^2$ queries.*

2.5 Quantized distributions and distance bounds

We now define the notion of quantized distributions, which will feature heavily in our estimation algorithm. Quantizing will allow us to apply "finite" algorithms over objects which are defined over infinite sets.

Definition 2.22 (ρ-quantized distribution). For $r \in \mathbb{N}$ and $\rho = 1/r \in (0,1)$, we say that a distribution μ over Ω is ρ-*quantized* if for any $w \in \Omega$ it holds that $\mu(w)$ is an integer multiple of ρ. That is, the measure takes the form $\mu : \Omega \to \{0, \rho, 2\rho, \ldots, 1\}$.

The following lemma is folklore. For completeness we provide it with a proof sketch.

Lemma 2.23. *Any distribution μ over a finite set Ω can be transformed to a ρ-quantized distribution μ' so that $\mu'(x) \in \{\rho\lceil\mu(x)/\rho\rceil, \rho\lfloor\mu(x)/\rho\rfloor\}$ for every $x \in \Omega$. In particular $\mu'(x) = \mu(x)$ whenever $\mu(x) \in \{0, 1\}$, $|\mu(x) - \mu'(x)| < \rho$ for all $x \in \Omega$, and $d_{\mathrm{TV}}(\mu, \mu') < \frac{\rho}{2} \cdot |\Omega|$.*

Proof: We construct μ' by finding for every $x \in \Omega$ a choice of $\mu'(x) \in \{\rho\lceil\mu(x)/\rho\rceil, \rho\lfloor\mu(x)/\rho\rfloor\}$ so that μ' will be a distribution, i.e. $\sum_{x \in \Omega} \mu'(x) = 1$. To produce this choice, we set an order $\Omega = \{x_1, \ldots, x_r\}$ where $r = |\Omega|$, and then inductively choose $\mu'(x_i) = \rho\lceil\mu(x_i)/\rho\rceil$ if $\sum_{j \in [i-1]} \mu'(x_j) \leq \sum_{j \in [i-1]} \mu(x_j)$ and $\mu'(x_i) = \rho\lfloor\mu(x_i)/\rho\rfloor$ otherwise. It is not hard to see that this way we maintain the invariant $\sum_{j \in [i]} \mu(x_j) - \rho < \sum_{j \in [i]} \mu'(x_j) < \sum_{j \in [i]} \mu(x_j) + \rho$, and in particular $\sum_{j \in [r]} \mu'(x) = 1$ since it is an integer multiple of ρ that lies strictly between $1 - \rho$ and $1 + \rho$. ∎

We continue with more definitions and results regarding approximate distributions, mainly centered around the quantization notion.

Definition 2.24 (ρ-legitimate approximation). Fix $t \in [0,1]^A$ and $\rho \in (0,1)$. A vector $\tilde{t} \in [0,1]^A$ is a ρ-*legitimate approximation* of t if $\|\tilde{t} - t\|_\infty \leq \rho$.

Observation 2.25. *Let η be a distribution over a set A. For any vector $t \in [0,1]^A$ we have a ρ-legitimate approximation $\tilde{t}$ satisfies $d_{\ell_1}^\eta(\tilde{t}, t) \leq \rho$.*

Proof: By definition of the η-weighted ℓ_1 distance $d_{\ell_1}^\eta(\tilde{t}, t) = \sum_{a \in A} \eta(a) \cdot |\tilde{t}(a) - t(a)| \leq \rho$. ∎

We will need to quantize not only the probabilities, but also the probability space itself (if it were not finite to begin with).

Definition 2.26. For $\rho \in (0,1)$ and a distribution Λ over $[0,1]^A$, we define Λ_ρ to be a distribution over $\{0, \rho, 2\rho, \ldots, 1\}^A$ obtained in the following way. Sample t according to Λ and round each entry in t to the nearest multiple of ρ to obtain a ρ-legitimate approximation $\tilde{t}$ of t.

Now we have the following lemma which connects these two definitions.

Lemma 2.27. *Let $\rho \in (0,1)$ and fix a distribution Λ over $[0,1]^A$. Then for any distribution η over A, $d_{\mathrm{EM}}^\eta(\Lambda, \Lambda_\rho) \leq \rho/2$.*

Proof: We consider the following transfer function $T : [0,1]^A \to [0,1]^A$ from Λ to Λ_ρ:

$$T(x,y) \overset{\mathrm{def}}{=} \begin{cases} \Lambda(x) & \text{if } y = \tilde{x} \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{x}$ denotes the vector obtained from x by rounding the entries to the nearest multiple of ρ, which is at most $\rho/2$ away from the unrounded value. From Observation 2.25 we have that whenever $T(x,y) \neq 0$ we have $d_{\ell_1}^\eta(x,y) \leq \rho/2$. Therefore,

$$d_{\mathrm{EM}}^\eta(\Lambda, \Lambda_\rho) \leq \sum_{x,y \in [0,1]^A} T(x,y) \cdot d_{\ell_1}^\eta(x,y) \leq \frac{\rho}{2} \sum_{x,y \in [0,1]^A} T(x,y) \leq \rho/2.$$

As a sort of a summary of the above, we now formulate a "quantize everything" lemma.

Lemma 2.28. *Fix $\rho = 1/r$ for some $r \in \mathbb{N}$, and let Λ be a distribution over $[0,1]^A$ for some finite set A of size ℓ. There exists a $\rho/(r+1)^\ell$-quantized distribution Λ' over $\{0, \rho, 2\rho, \ldots, 1\}^A$ such that $d_{\mathrm{EM}}^\eta(\Lambda', \Lambda) < \rho$.*

Proof: We use Lemma 2.27 to obtain a distribution Λ_ρ over $\Omega = \{0, \rho, 2\rho, \ldots, 1\}^A$ such that $d_{\mathrm{EM}}^\eta(\Lambda, \Lambda_\rho) \leq \rho/2$, and apply Lemma 2.23 using $\rho' = \rho/(r+1)^\ell = \rho/|\Omega|$ to obtain a ρ'-quantized distribution Λ' over $\{0, \rho, 2\rho, \ldots, 1\}^A$, such that $d_{\mathrm{TV}}(\Lambda_\rho, \Lambda') \leq \rho/2$. Overall, using the triangle inequality, we have:
$$d_{\mathrm{EM}}^\eta(\Lambda', \Lambda) \leq d_{\mathrm{EM}}^\eta(\Lambda', \Lambda_\rho) + d_{\mathrm{EM}}^\eta(\Lambda_\rho, \Lambda) \leq \rho.$$

∎

We conclude with a lemma that allows us to move to guarantee a small change in the weighted Earth Mover distance when we need to use an *approximation* of the weight distribution η.

Lemma 2.29. *Let η and η' be two distributions over A satisfying $d_{\mathrm{TV}}(\eta, \eta') \leq \rho$, and let Λ and Υ be two discrete distributions over $[0,1]^A$. Then, $d_{\mathrm{EM}}^{\eta'}(\Lambda, \Upsilon) \leq d_{\mathrm{EM}}^\eta(\Lambda, \Upsilon) + 2\rho$.*

Proof: Let T denote an optimal transfer function between Λ and Υ with respect to $d_{\mathrm{EM}}^\eta(\Lambda, \Upsilon)$ (which might be sub-optimal with respect for $d_{\mathrm{EM}}^{\eta'}(\Lambda, \Upsilon)$). Using also the fact that $|t(a) - t'(a)| \leq 1$ for any $t, t' \in [0,1]^A$ and $a \in A$ we obtain:

$$
\begin{aligned}
d_{\mathrm{EM}}^{\eta'}(\Lambda, \Upsilon) - d_{\mathrm{EM}}^\eta(\Lambda, \Upsilon) &\leq \mathop{\mathbf{E}}_{(\boldsymbol{t}, \boldsymbol{t}') \sim T}\left[\mathop{\mathbf{E}}_{\boldsymbol{a} \sim \eta'}|\boldsymbol{t}(\boldsymbol{a}) - \boldsymbol{t}'(\boldsymbol{a})|\right] - \mathop{\mathbf{E}}_{(\boldsymbol{t}, \boldsymbol{t}') \sim T}\left[\mathop{\mathbf{E}}_{\boldsymbol{a} \sim \eta}|\boldsymbol{t}(\boldsymbol{a}) - \boldsymbol{t}'(\boldsymbol{a})|\right] \\
&= \mathop{\mathbf{E}}_{(\boldsymbol{t}, \boldsymbol{t}') \sim T}\left[\sum_{a \in A}(\eta'(a) - \eta(a)) \cdot |\boldsymbol{t}(\boldsymbol{a}) - \boldsymbol{t}'(\boldsymbol{a})|\right] \\
&\leq \mathop{\mathbf{E}}_{(\boldsymbol{t}, \boldsymbol{t}') \sim T}\left[\sum_{a \in A}|\eta'(a) - \eta(a)| \cdot |\boldsymbol{t}(\boldsymbol{a}) - \boldsymbol{t}'(\boldsymbol{a})|\right] \\
&\leq \sum_{a \in A}|\eta'(a) - \eta(a)| \leq 2d_{\mathrm{TV}}(\eta, \eta') \leq 2\rho.
\end{aligned}
$$

∎

2.6 Some useful probabilistic inequalities

We will use some well-known large deviation inequalities.

Lemma 2.30 (Multiplicative Chernoff bound, see [DP09]). *Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be independent random variables such that $\mathbf{X}_i \in [0,1]$ for every $i \in [n]$. For $\mathbf{X} = \sum_{i=1}^n \mathbf{X}_i$, the following holds for any $0 \leq \delta \leq 1$.*
$$\Pr[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| \geq \delta\,\mathbf{E}[\mathbf{X}]] \leq 2\exp\left(-\mathbf{E}[\mathbf{X}]\delta^2/3\right).$$

Lemma 2.31 (Additive Chernoff bound, see [DP09]). *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be independent random variables such that* $\mathbf{X}_i \in [0, 1]$ *for every* $i \in [n]$. *For* $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$ *and* $a \leq \mathbf{E}[\mathbf{X}] \leq b$, *the following hold for any* $\delta > 0$.

(i) $\mathbf{Pr}\left[\mathbf{X} \geq b + \delta\right] \leq \exp\left(-2\delta^2/n\right)$.

(ii) $\mathbf{Pr}\left[\mathbf{X} \leq a - \delta\right] \leq \exp\left(-2\delta^2/n\right)$.

Lemma 2.32 (Hoeffding's Inequality, see [DP09]). *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be independent random variables such that* $a_i \leq \mathbf{X}_i \leq b_i$ *and* $\mathbf{X} = \sum_{i=1}^{n} \mathbf{X}_i$. *Then, for all* $\delta > 0$,

$$\mathbf{Pr}\left[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| \geq \delta\right] \leq 2 \exp\left(-2\delta^2/\sum_{i=1}^{n}(b_i - a_i)^2\right).$$

Lemma 2.33 (Hoeffding's Inequality for sampling without replacement [Hoe94]). *Let* n *and* m *be two integers such that* $1 \leq n \leq m$, *and* $x_1, \ldots, x_m$ *be real numbers, with* $a \leq x_i \leq b$ *for every* $i \in [m]$. *Suppose that* $\mathbf{I}$ *is a set that is drawn uniformly from all subsets of* $[m]$ *of size* n, *and let* $\mathbf{X} = \sum_{i \in \mathbf{I}} x_i$. *Then, for all* $\delta > 0$,

$$\mathbf{Pr}\left[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| \geq \delta\right] \leq 2 \exp\left(-2\delta^2/n \cdot (b - a)^2\right).$$

We will also use some second moment lower bounds.

Lemma 2.34 (Cauchy-Schwartz inequality). *Given any distribution with two random variables* $\mathbf{X}$ *and* $\mathbf{Z}$,

$$\mathbf{\underset{Z}{E}}\left[\mathbf{E}[\mathbf{X}^2 \mid \mathbf{Z}]\right] \geq \mathbf{\underset{Z}{E}}\left[\mathbf{E}[\mathbf{X} \mid \mathbf{Z}]^2\right].$$

This is a direct implication of $\mathbf{E}[\mathbf{X}^2] - (\mathbf{E}[\mathbf{X}])^2 = \mathbf{E}[(\mathbf{X} - \mathbf{E}[\mathbf{X}])^2] \geq 0$ *when considered over any possible conditioning of the probability space to a value of* $\mathbf{Z}$.

Observation 2.35 (Defect form Cauchy-Schwartz inequality). *If* $\mathbf{X}$ *is a random variable satisfying* $\mathbf{Pr}[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| \geq \alpha] \geq \beta$, *then*

$$\mathbf{E}[\mathbf{X}^2] - (\mathbf{E}[\mathbf{X}])^2 = \mathbf{E}[(\mathbf{X} - \mathbf{E}[\mathbf{X}])^2] \geq \alpha^2 \beta.$$

In addition, we will use the following well-known bound.

Lemma 2.36 (Reverse Markov's inequality). *For any* $\gamma \in (0, 1)$, *let* $\mathbf{X}$ *be a random variable in* $[-1, 1]$ *with* $\mathbf{E}[\mathbf{X}] \geq \gamma$. *Then* $\mathbf{Pr}[\mathbf{X} \geq \gamma/2] \geq \gamma/2$.

3 Detailing, variable types and robustness

In this section we define the main objects that we use throughout this work.

3.1 Detailing and refinements

Our most basic definition is that of a detailing of a distribution μ. Essentially it is a possible "deconstruction" providing additional information to aid with the analysis of μ, in analogy to the notion of a vertex partition of a graph.

Definition 3.1 (Detailing). Let μ be a distribution over Ω and A be a set. A *detailing* ξ of μ with respect to A is a distribution over $\Omega \times A$ that satisfies $\xi|_1 = \mu$. $|A|$ is referred to as the *length* of the detailing. When $|A| = 1$, we call ξ the *trivial detailing* and identify it with μ itself.

The following immediate observation is in fact the motivation for Definition 2.5.

Observation 3.2. *If ν is a distribution over A and η is a distribution over $A \times B$, then the adjustment $\nu \rhd \eta$ of η to ν is in particular a detailing of ν with respect to B.*

We will also define ρ-quantized detailings, even when the original distribution is not quantized. Such detailings are not quantized by themselves, only their "detailing portion" is quantized.

Definition 3.3 (Quantized detailing). Let μ be a distribution over Ω, A be a set and ξ be a detailing of μ with respect to A. For $\rho \in (0, 1)$ for which $1/\rho \in \mathbb{N}$, we say that ξ is *ρ-quantized* if for every $x \in \Omega$ for which $\xi|_1(x) > 0$, the conditional distribution $\xi|^{1:x}$ is ρ-quantized.

The following is an immediate consequence of Lemma 2.23.

Observation 3.4. *For a distribution μ on a finite set Ω, any detailing ξ of μ with respect to a finite A can be transformed to a ρ-quantized detailing ξ', so that $\xi'|^{1:x}(a) \in \{\rho\lceil\xi|^{1:x}(a)/\rho\rceil, \rho\lfloor\xi|^{1:x}(a)/\rho\rfloor\}$ for every $x \in \Omega$ and $a \in A$. In particular $\xi'(x, a) = \xi(x, a)$ whenever $\xi'|^{1:x}(a) \in \{0, 1\}$ and $d_{\mathrm{TV}}(\xi, \xi') < \frac{\rho}{2} \cdot |A|$.*

Proof: We will use Lemma 2.23 over $\xi|^{1:x}$ for every $x \in \Omega$ with $\xi|_1(x) > 0$. ■

Any given detailing of a distribution can be "detailed further", resulting in a refinement.

Definition 3.5 (Refinement of a detailing). Let μ be a distribution defined over Ω and ξ be a detailing of μ with respect to A. A *refinement* of ξ by a set B is a detailing ξ' of ξ with respect to B. As there is a natural bijection between $\Omega \times (A \times B)$ and $(\Omega \times A) \times B$, ξ' can and will be considered also as a detailing of μ with respect to $A \times B$.

Specific to our investigation, we also need a definition for a "do-nothing" detailing (or refinement), to which we refer as flat.

Definition 3.6 (Flat detailing). Let μ be a distribution over Ω and A be a set. A detailing ξ of μ with respect to A is said to be a *flat detailing* if it is a product distribution over $\Omega \times A$, i.e. $\xi(x, a) = \mu(x) \cdot \xi|_2(a)$ for every $x \in \Omega, a \in A$.

Definition 3.7 (Flat refinement of a detailing). Let μ be a distribution over $\Omega = \{0, 1\}^n$ and ξ be a detailing of μ with respect to A. Consider a refinement ξ' of ξ with respect to B, i.e., a detailing of μ with respect to $A \times B$ for which $\xi'|_{1,2} = \xi$. Such a ξ' is said to be a *flat refinement* if it is a flat detailing of ξ.

A detailing will usually have a continuum of possible refinements, even with respect to a set of size 2. Moreover, there is no good way to sample from a detailing or a refinement in its general form when we only have access to samples from μ, and even then we only have query access to the samples from μ. A specific way to define a detailing which is tangible for a testing algorithm is to "partition" μ by the restriction of the samples to a specific set of variables.

Definition 3.8 (A refinement by a set of variables U). Let μ be a distribution over $\Omega = \{0,1\}^n$ and ξ be a detailing of μ with respect to a set A. For a subset $U \subseteq [n]$, the *refinement ξ^U of ξ by the set U* is the distribution over $\Omega \times (A \times \{0,1\}^{|U|})$ that satisfies the following:

1. $\xi^U(x, a, v) = 0$ if $x_U \neq v$, and

2. $\xi^U(x, a, v) = \xi(x, a)$ if $x_U = v$.

Note that the length of the detailing ξ^U is $|A| \cdot 2^{|U|}$. When ξ is the trivial detailing (i.e. $\xi = \mu$) we denote the refinement by U as μ^U, and also call it the *detailing of μ defined by the variable set U*.

3.2 Type distributions

When trying to analyze a distribution by its detailing with respect to a set A, a major role is given to its "weights" as given by its projection over A.

Definition 3.9 (Weight distribution of a detailing). Let μ be a distribution defined over Ω, and let ξ be a detailing of μ with respect to a set A. The *weight distribution of ξ* is defined as $\xi|_2$.

We sometimes go the other way and define a flat refinement of a detailing ξ by its weight distribution (which should "conform" to the weight distribution of ξ).

Definition 3.10 (Flat refinement by weights). Given a detailing ξ of μ over A, and a detailing η of $\xi|_2$ over B (that is, a distribution over $A \times B$ for which $\eta|_1 = \xi|_2$, thus "extending" the weight distribution of ξ), we denote by $\xi_{\langle \eta \rangle}$ the specific flat refinement of ξ defined by $\xi_{\langle \eta \rangle}(x, a, b) = \xi(x, a) \cdot \eta|_2^{1:a}(b) = \xi|_1^{2:a}(x) \cdot \eta(a, b)$. In other words, $\xi_{\langle \eta \rangle} = \xi \bowtie \eta$.

We now define the other important aspect of a detailing of a distribution over $\{0,1\}^n$, the "type" of a variable $i \in [n]$ as defined by its distribution in the component $\xi|_1^{2:a}$ for every $a \in A$.

Definition 3.11 (Variable types of a detailing). Let μ be a distribution of $\{0,1\}^n$. Let us consider a detailing ξ of μ with respect to A. The *type t_i of a variable $i \in [n]$* with respect to ξ, is a vector in $[0,1]^A$ defined by

$$t_i \overset{\text{def}}{=} \left\langle \Pr_{(\boldsymbol{x},a) \sim \xi|^{\mu \times \{a\}}} [\boldsymbol{x}_i = 1] \right\rangle_{a \in A}.$$

For any $a \in A$ for which $\xi|_2(a) = 0$, the value $t_i(a)$ is set arbitrarily (and will not affect any computation in the following).

Definition 3.12 (Type distribution). Let ξ be a detailing of μ with respect to A. The *type distribution* of ξ is denoted by Λ and defined over $[0,1]^A$, where a sample from Λ is obtained as follows: Choose $i \in [n]$ uniformly at random and report its type t_i.

We also look at the types that we would expect from a flat refinement of a detailing.

Definition 3.13 (Flat extension). Given a vector $t \in [0,1]^A$ and a set B, its *flat extension* with respect to B is the vector $t' \in [0,1]^{A \times B}$ defined by $t'(a,b) = t(a)$ for all $a \in A$ and $b \in B$. Given a finitely supported distribution Λ over $[0,1]^A$, its *flat extension* with respect to B, denoted by $\Lambda_{\langle B \rangle}$, is correspondingly defined as the distribution for which $\Lambda_{\langle B \rangle}(t') = \Lambda(t)$ for any $t \in [0,1]^A$ and $t' \in [0,1]^{A \times B}$ which is the flat extension of t.

Observation 3.14. *If ξ is a detailing of A, and ξ' is any flat refinement of ξ with respect to B, then for any i, the type t'_i of any variable i with respect to ξ' is exactly the flat extension of the type t_i of i by ξ with respect to B (not counting pairs (a,b) for which $\xi'|_{2,3}(a,b) = 0$, for which $t'_i(a,b)$ is anyway arbitrarily defined).*

Respectively, the type distribution of such an ξ' is the flat extension with respect to B of the type distribution of ξ.

While usually converting the multi-set of variables' types to a distribution as per Definition 3.12, we sometimes consider the other direction of starting with a distribution and "implementing" it as an ordered sequence of elements. We formulate the following with respect to a general space Ω, and then provide some lemmas for a corresponding general Earth Mover distance, but here we will almost always use the space $[0,1]^A$ for some A and an η-weighted ℓ_1 metric for some η, as befits distributions over variable types.

Definition 3.15 (Implementation). Let Λ be a discrete distribution over Ω. An *implementation* of Λ is a function $h : [n] \to \Omega$, so that $|\{i \in [n] : h(i) = t\}| = n \cdot \Lambda(t)$ for all $t \in \Omega$.

Given a detailing ξ of some μ over $\{0,1\}^n$ admitting type distribution Λ, the *implementation of Λ demonstrated by the detailing ξ* is the map h defined by $h(i) = t_i$, where t_i is the type of i under the detailing ξ.

Note that an implementation of Λ is equivalent to a uniform distribution π over n tuples $(i, w) \in [n] \times \Omega$, where every $i \in [n]$ appears in exactly one of the pairs and $\Lambda(w) = \mathbf{Pr}_\pi[[n] \times \{w\}] = \pi|_2(w)$.

The proof of the following lemma is obtained using linear programming.

Lemma 3.16 (Follows from the proof of Claim 6.5 in [CFG$^+$23]). *Fix a space Ω with a distance measure d. If two finitely supported distributions Λ and Λ' over Ω are $\frac{1}{n}$-quantized, then there exists a $\frac{1}{n}$-quantized transfer function T from Λ to Λ' which is optimal with respect to $d_{\mathrm{EM}}(\Lambda, \Lambda')$, that is,*

$$d_{\mathrm{EM}}(\Lambda, \Lambda') = \mathop{\mathbf{E}}_{(\boldsymbol{t}, \boldsymbol{t'}) \sim T} \left[d(\boldsymbol{t}, \boldsymbol{t'}) \right]$$

The following "satisfiable pigeon-hole principal" is trivial.

Observation 3.17. *If L is a multiset of members from $X \times Y$ and N is a multiset of members from $Y \times Z$, and the multisets $\langle b : (a,b) \in L \rangle$ and $\langle b : (b,c) \in N \rangle$ are identical, then there is a multiset J of members from $X \times Y \times Z$ for which $\langle (a,b) : (a,b,c) \in J \rangle = L$ and $\langle (b,c) : (a,b,c) \in J \rangle = N$.*

Using Lemma 3.16 and Observation 3.17 the following lemma is immediate, by identifying the $\frac{1}{n}$-quantized distributions with the corresponding n element multisets, and using the fact that an implementation is a $\frac{1}{n}$-quantized distribution over $[n] \times \Omega$, whose restriction to the first coordinate is uniform over $[n]$.

Lemma 3.18. *Fix a space Ω with a distance mesure d. If Λ and Λ' are $\frac{1}{n}$-quantized, than any implementation h of Λ can be extended to an implementation H of a transfer distribution κ between Λ and Λ' such that*

$$d_{\mathrm{EM}}(\Lambda, \Lambda') = \mathop{\mathbf{E}}_{(\boldsymbol{t}, \boldsymbol{t}') \sim \kappa} \left[d(\boldsymbol{t}, \boldsymbol{t}') \right] = \mathop{\mathbf{E}}_{\boldsymbol{i} \sim [n]} \left[d(((H(\boldsymbol{i}))_1, (H(\boldsymbol{i}))_2)) \right]$$

We end by noting that the above lemma has an easy "converse".

Observation 3.19. *Let Λ and Λ' be distributions over some Ω with distance measure d, let $h : [n] \to \Omega$ be an implementation of Λ and $h' : [n] \to \Omega$ be an implementation of Λ'. Then,*

$$d_{\mathrm{EM}}(\Lambda, \Lambda') \leq \mathop{\mathbf{E}}_{\boldsymbol{i} \sim [n]} \left[d(h(\boldsymbol{i}), h'(\boldsymbol{i})) \right].$$

Proof: By the definition of $d_{\mathrm{EM}}(\Lambda, \Lambda')$, it is enough to construct a transfer distribution T between Λ and Λ' for which $\mathbf{E}_{(\boldsymbol{t}, \boldsymbol{t}') \sim T} \left[d(\boldsymbol{t}, \boldsymbol{t}') \right] \leq \mathbf{E}_{\boldsymbol{i} \sim [n]} \left[d(h(\boldsymbol{i}), h'(\boldsymbol{i})) \right]$. To construct such a T, we define a draw $(\boldsymbol{t}, \boldsymbol{t}') \sim T$ as the result of uniformly drawing $\boldsymbol{i} \sim [n]$ and setting $(\boldsymbol{t}, \boldsymbol{t}') = (h(\boldsymbol{i}), h'(\boldsymbol{i}))$. $\blacksquare$

3.3 Notions of robustness

Definition 3.20 (Index of a detailing)**.** Let ξ be a detailing of μ with respect to A. The *index* of ξ is defined as

$$\mathrm{Ind}(\xi) \stackrel{\mathrm{def}}{=} \mathop{\mathbf{E}}_{\boldsymbol{i} \sim [n]} \left[\mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\mathbf{Pr}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}} [\boldsymbol{x}_i = 1]^2 \right] \right]$$

Note that for any detailing ξ, $0 \leq \mathrm{Ind}(\xi) \leq 1$.

We first define the robustness of a detailing in its general form.

Definition 3.21 (Robust detailing)**.** Let $\delta \in (0, 1)$ and $\ell \in \mathbb{N}$ be constants. A detailing ξ of μ with respect to A is called (δ, ℓ)-*robust* if there exists no refinement ξ' of ξ with respect to a set B such that $|B| \leq \ell$ and $\mathrm{Ind}(\xi') \geq \mathrm{Ind}(\xi) + \delta$.

The second type of robustness concerns only refinements obtained by a set of variables, and even then puts its requirement over most (rather than all) candidate variable sets.

Definition 3.22 (Weakly robust detailing)**.** Let $\delta \in (0, 1)$ and $k \in \mathbb{N}$ be constants. A detailing ξ is said to be (δ, k)-*weakly robust* if for at least a $1 - \delta$ fraction of the sets $U \subseteq [n]$ of size at most k, it holds that $\mathrm{Ind}(\xi^U) < \mathrm{Ind}(\xi) + \delta$.

A useful feature of the way robustness is defined, is that the existence of a robust refinement of a given detailing follows almost immediately from the definition. We state here the variant of such a lemma that will be useful to us later on.

Lemma 3.23. *For any k and δ, if ξ is any detailing of μ with respect to a set A, then there exists a set U of size at most k/δ such that ξ^U is (δ, k)-robust.*

Proof: We consider the following sequence of sets $U_0, U_1, \ldots, U_r$ defined by the following inductive process: We start with the basis $U_0 = \emptyset$, and given U_i we check whether there exists a set $V_i \subset [n]$ of size at most k so that $\mathrm{Ind}(\xi^{U_i \cup V_i}) \geq \mathrm{Ind}(\xi^{U_i}) + \delta$. If there exists such a set, we set $U_{i+1} = U_i \cup V_i$ and continue this process for U_{i+1}. Otherwise, we set $r = i$, $U = U_i$ and terminate the process.

By definition, once this process is terminated, we obtain a (δ, k)-weakly robust detailing ξ^U (in fact the termination condition is slightly stronger than the condition of Definition 3.22). Additionally, noting that for every $i \leq r$ we have $\mathrm{Ind}(\xi^{U_i}) \geq \mathrm{Ind}(\xi) + i \cdot \delta$ and $|U_i| \leq ik$ by induction, and that $\mathrm{Ind}(\xi^W) \in [0, 1]$ by definition, for any set $W \subseteq [n]$, we conclude that $r \leq 1/\delta$ and so $|U| \leq rk \leq k/\delta$, as required. $\blacksquare$

We now define what it means for a detailing to be (ϵ, k)-good, a notion that will provide us a degree of predictability with respect to testing algorithms.

Definition 3.24 (ϵ-independent tuple). Fix $\epsilon \in (0, 1)$ and let μ be a distribution over $\{0, 1\}^n$, and $i_1, \ldots, i_k \in [n]$ be a set of integers. We say that the k-tuple $(i_1, \ldots, i_k)$ an ϵ-*independent tuple* with respect to μ if $d_{\mathrm{TV}}\left(\mu|_{\{i_1, \ldots, i_k\}}, \prod_{\ell=1}^{k} \mu|_{i_\ell}\right) \leq \epsilon$.

Definition 3.25 ((ϵ, k)-good detailing). Let μ be a distribution supported on $\{0, 1\}^n$ and let ξ be a detailing of μ over some A. Let $J \subseteq A$ be the set of elements where $a \in J$ if at least $(1 - \epsilon)n^k$ of the k-tuples in $[n]$ are ϵ-independent with respect to $\xi|_1^{2:a}$. We call the detailing (ϵ, k)-*good* if $\mathbf{Pr}_{a \sim \xi|_2}[a \in J] \geq 1 - \epsilon$, and refer to J as the *good set*.

3.4 From variable weak robustness to goodness

We prove here that variable weak robustness implies goodness, as stated in the following lemma.

Lemma 3.26. *If a detailing ξ is $(\frac{\epsilon^6}{64k^3}, k - 1)$-weakly robust, then it is also (ϵ, k)-good.*

Before we prove the lemma, we set the stage with some technical tools and definitions. In the following, for a distribution μ supported on $\{0, 1\}^n$, given a tuple $\alpha = (i_1, \ldots, i_k) \in [n]^k$ and an assignment $v = (v_{i_1}, \ldots, v_{i_k}) \in \{0, 1\}^k$ we let $\mu|^{i_1:v_{i_1}, \ldots, i_k:v_{i_k}}$ denote the conditioned distribution $\mu|^{x_\alpha = v}$. Note that this is not a new definition but just a new use for Definition 2.3.

Lemma 3.27. *If the tuple $\{i_1, \ldots, i_k\}$ is not ϵ-independent, then there exists an index $1 \leq \ell \leq k$ and a set $V \subseteq \{0, 1\}^{k-1}$, such that $\mathbf{Pr}_{\mu|_{\{i_1, \ldots, i_{\ell-1}, i_{\ell+1}, \ldots, i_k\}}}[V] \geq \epsilon/2k$, and for every member $v = (v_{i_1}, \ldots, v_{i_{\ell-1}}, v_{i_{\ell+1}}, \ldots, v_{i_k}) \in V$ we have $\left|\mu|_{i_\ell}(1) - \mu|_{i_\ell}^{i_1:v_1, \ldots, i_{\ell-1}:v_{\ell-1}, i_{\ell+1}:v_{\ell+1}, \ldots, i_k:v_k}(1)\right| \geq \epsilon/2k$.*

The proof of Lemma 3.27 follows from the following one.

Lemma 3.28. *Let μ be a distribution over $\{0, 1\}^n$. Then*

$$d_{\mathrm{TV}}\left(\mu|_{\{i_1, \ldots, i_k\}}, \prod_{j=1}^{k} \mu|_{i_j}\right) \leq \sum_{j=1}^{\ell} \mathop{\mathbf{E}}_{x \sim \mu}\left[\left|\mu|_{i_\ell}^{i_1:x_{i_1}, \ldots, i_{\ell-1}:x_{i_{\ell-1}}}(1) - \mu|_{i_\ell}(1)\right|\right]$$

Proof: We use Observation 2.10, and prove the statement by constructing an appropriate coupling (transfer distribution) T which demonstrates this distance bound (with respect to the Kronecker norm) between $\nu \stackrel{\text{def}}{=} \mu|_{\{i_1, \ldots, i_k\}}$ and $\tau \stackrel{\text{def}}{=} \prod_{j=1}^{k} \mu|_{i_j}$. To draw $(x, y) \sim T$ (where $(x, y) \in (\{0, 1\}^k)^2$),

18

for $\ell \in [k]$ we draw $\boldsymbol{x}_\ell$ and $\boldsymbol{y}_\ell$ inductively considering the values of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\ell-1}$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{\ell-1}$. The base case (before we draw any bits) is trivial.

To draw $\boldsymbol{x}_\ell$ and $\boldsymbol{y}_\ell$ we first calculate

$$\alpha_\ell = \Pr_{\boldsymbol{z} \sim \mu|^{i_1:\boldsymbol{x}_1,\ldots,i_{\ell-1}:\boldsymbol{x}_{\ell-1}}} [\boldsymbol{z}_{j_\ell} = 1] = \mu|^{i_1:\boldsymbol{x}_1,\ldots,i_{\ell-1}:\boldsymbol{x}_{\ell-1}}_{i_\ell}(1)$$

and $\beta_\ell = \mu|_{i_\ell}(1)$. Then if $\alpha_\ell \geq \beta_\ell$ we set $(\boldsymbol{x}_\ell, \boldsymbol{y}_\ell) = (1,1)$ with probability β_ℓ, $(\boldsymbol{x}_\ell, \boldsymbol{y}_\ell) = (1,0)$ with probability $\alpha_\ell - \beta_\ell$, and $(\boldsymbol{x}_\ell, \boldsymbol{y}_\ell) = (0,0)$ with probability $1 - \alpha_\ell$. If $\alpha_\ell \leq \beta_\ell$ then we analogously set $(\boldsymbol{x}_\ell, \boldsymbol{y}_\ell) = (1,1)$ with probability α_ℓ, $(\boldsymbol{x}_\ell, \boldsymbol{y}_\ell) = (0,1)$ with probability $\beta_\ell - \alpha_\ell$, and $(\boldsymbol{x}_\ell, \boldsymbol{y}_\ell) = (0,0)$ with probability $1 - \beta_\ell$.

It is not hard to see that this is indeed a coupling between ν and τ. Also, for every ℓ clearly

$$\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim T}[\boldsymbol{x}_\ell \neq \boldsymbol{y}_\ell] = \mathbf{E}_{\boldsymbol{x} \sim \mu}[|\alpha_\ell - \beta_\ell|] = \mathbf{E}_{\boldsymbol{x} \sim \mu}\left[\left|\mu|^{i_1:\boldsymbol{x}_{i_1},\ldots,i_{\ell-1}:\boldsymbol{x}_{j_{\ell-1}}}_{i_\ell}(1) - \mu|_{i_\ell}(1)\right|\right]$$

and hence by a union bound

$$\Pr_{(\boldsymbol{x},\boldsymbol{y}) \sim T}[\boldsymbol{x} \neq \boldsymbol{y}] \leq \sum_{j=1}^{\ell} \mathbf{E}_{\boldsymbol{x} \sim \mu}\left[\left|\mu|^{i_1:\boldsymbol{x}_{i_1},\ldots,i_{\ell-1}:\boldsymbol{x}_{j_{\ell-1}}}_{i_\ell}(1) - \mu|_{i_\ell}(1)\right|\right]$$

completing the proof. ∎

Proof of Lemma 3.27: Assume that the conclusion of the lemma does not hold. We define for all $\ell \in [k]$ the set $V_\ell \subseteq \{0,1\}^{k-1}$ of all v for which $\left|\mu|_{i_\ell}(1) - \mu|^{i_1:v_1,\ldots,i_{\ell-1}:v_{\ell-1},i_{\ell+1}:v_{\ell+1},\ldots,i_k:v_k}_{i_\ell}(1)\right| \geq \epsilon/2k$. We then note that $\delta_\ell = \mathbf{E}_{\boldsymbol{x} \sim \mu}\left[\left|\mu|^{i_1:\boldsymbol{x}_{i_1},\ldots,i_{\ell-1}:\boldsymbol{x}_{j_{\ell-1}}}_{i_\ell}(1) - \mu|_{i_\ell}(1)\right|\right]$ satisfies

$$\delta_\ell = \mathbf{E}_{\boldsymbol{x} \sim \mu}\left[\left|\mathbf{E}_{\boldsymbol{z} \sim \mu}\left[\mu|^{i_1:\boldsymbol{x}_{i_1},\ldots,i_{\ell-1}:\boldsymbol{x}_{j_{\ell-1}},i_{\ell+1}:\boldsymbol{z}_{i_{\ell+1}},\ldots,i_k:\boldsymbol{z}_k}_{i_\ell}(1) - \mu|_{i_\ell}(1)\right]\right|\right]$$

$$\leq \mathbf{E}_{\boldsymbol{x} \sim \mu}\left[\left|\mu|^{i_1:\boldsymbol{x}_{i_1},\ldots,i_{\ell-1}:\boldsymbol{x}_{j_{\ell-1}},i_{\ell+1}:\boldsymbol{x}_{i_{\ell+1}},\ldots,i_k:\boldsymbol{x}_k}_{i_\ell}(1) - \mu|_{i_\ell}(1)\right|\right]$$

$$\leq \Pr_{\mu|_{\{i_1,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_k\}}}[V_\ell] \cdot 1 + \Pr_{\mu|_{\{i_1,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_k\}}}[\{0,1\}^{k-1} \setminus V_\ell] \cdot \frac{\epsilon}{2k} = \frac{\epsilon}{k}$$

Then Lemma 3.28 would have implied that $d_{\mathrm{TV}}(\mu|_{\{i_1,\ldots,i_k\}}, \prod_{\ell=1}^{k} \mu|_{i_\ell}) \leq \sum_{\ell \in [k]} \delta_\ell \leq \epsilon$, in contradiction to our premise. ∎

Definition 3.29 (Pivot tuple). Fix $k \in [n]$, let μ be a distribution over $\{0,1\}^n$ and $\epsilon \in (0,1)$. We say that a $(k-1)$-tuple $(i_1, \ldots, i_{\ell-1}, i_{\ell+1}, \ldots, i_k) \in [n]^{k-1}$ is a *pivot tuple for i_ℓ with respect to μ* if the following holds. There exists a set $V \subseteq \{0,1\}^{k-1}$ such that $\Pr_{\mu|_{\{i_1,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_k\}}}[V] \geq \epsilon/2k$, and for every $v \in V$ we have $\left|\mu|_{i_\ell}(1) - \mu|^{i_1:v_1,\ldots,i_{\ell-1}:v_{\ell-1},i_{\ell+1}:v_{\ell+1},\ldots,i_k:v_k}_{i_\ell}(1)\right| \geq \epsilon/2k$.

We now prove, in two stages, that if a detailing is not (ϵ, k)-good, then there exist many tuples that are pivot over much of the weight for many variables.

Observation 3.30. *Fix $k \in [n]$, let μ be a distribution over $\{0,1\}^n$ and $\epsilon \in (0,1)$. If there exists a set $\mathcal{S} \in [n]^k$ of at least ϵn^k tuples that are not ϵ-independent, then there is a set $\mathcal{S}' \subseteq [n]^{k-1}$ of at least $\frac{\epsilon}{2k} \cdot n^{k-1}$ tuples such that each of them is a pivot tuple for at least $\frac{\epsilon}{2k} \cdot n$ variables outside of it.*

Proof: Note that Lemma 3.27 immediately implies that for every $U \in \mathcal{S}$ there exists $i_U \in U$ so that $U \setminus \{i_U\}$ is a pivot tuple with respect to i_U. By the premise of the observation, if we uniformly pick a tuple $U' \in [n]^{k-1}$ and $i \in [n]$, then with probability at least ϵ we will obtain that $U' \cup \{i\} \in \mathcal{S}$ (and also in particular $|U' \cup \{i\}| = k$), and so with probability at least ϵ/k we obtain that U' is a pivot tuple for i (since the probability for $i = i_{U' \cup \{i\}}$ conditioned on $U' \cup \{i\} = U \in \mathcal{S}$ is $1/k$).

A reverse Markov inequality then implies that there exists a set $\mathcal{S}'$ of at least $\frac{\epsilon}{2k} \cdot n^{k-1}$ tuples so that for every $U' \in \mathcal{S}'$ there exist at least $\frac{\epsilon}{2k} \cdot n$ many variables for which it is a pivot tuple. $\blacksquare$

Lemma 3.31. *If a detailing ξ of μ over A is not (ϵ, k)-good, then there exists a set $\mathcal{S} \subseteq [n]^{k-1}$ of at least $\frac{\epsilon}{4k} \cdot n^{k-1}$ tuples, so that for every $U \in \mathcal{S}$ there exists a set $K_U \subseteq A$ for which $\mathbf{Pr}_{\xi|_2}[K_U] \geq \frac{\epsilon^2}{4k}$, satisfying that for every $a \in K_U$ the tuple U is a pivot tuple for at least $\frac{\epsilon}{2k} \cdot n$ many variables with respect to $\xi|_1^{2:a}$.*

Proof: We let K be the set of all $a \in A$ for which $\xi|^{2:a}$ admits a set $\mathcal{S}_a \subseteq [n]^k$ of at least ϵn^k tuples which are not ϵ-independent. By the premise of ξ not being (ϵ, k)-good we have $\mathbf{Pr}_{\xi|_2}[K] \geq \epsilon$. By invoking Observation 3.30 for every $a \in K$, we obtain $\mathcal{S}'_a \subseteq [n]^{k-1}$ of size at least $\frac{\epsilon}{2k} \cdot n^{k-1}$ so that every $U \in \mathcal{S}'_a$ is a pivot tuple for at least $\frac{\epsilon}{2k} \cdot n$ many variables with respect to $\xi|^{2:a}$.

In particular, if we draw uniformly a tuple $U \in [n]^{k-1}$ and according to $\xi|_2^{2:K}$ a value $a \in K$, with probability at least $\frac{\epsilon}{2k}$ the tuple U is pivot tuple for at least $\frac{\epsilon}{2k} \cdot n$ many variables with respect to $\xi|^{2:a}$. To finalize, we use a reverse Markov inequality, and find a set $\mathcal{S} \subseteq [n]^{k-1}$ of at least $\frac{\epsilon}{4k} \cdot n^{k-1}$ many tuples, so that every $U \in \mathcal{S}$ satisfies the above assertion with respect to a set $K_U \subseteq K$ satisfying $\mathbf{Pr}_{\xi|_2^{2:K}}[K_U] \geq \frac{\epsilon}{4k}$ and hence $\mathbf{Pr}_{\xi|_2}[K_U] \geq \frac{\epsilon^2}{4k}$. $\blacksquare$

We are now ready to prove that a sufficient (weak) robustness of a detailing implies its goodness.

Proof of Lemma 3.26: As ξ is not (ϵ, k)-good, we use Lemma 3.31 to find the set $\mathcal{S} \subseteq [n]^{k-1}$ satisfying its conclusion, and also note the set $K_U \subseteq A$ promised by the lemma for every $U \in \mathcal{S}$. Additionally, for every $U \in \mathcal{S}$ and $a \in K_U$ we note the set $I_{U,a} \subseteq n$ of (at least $\frac{\epsilon}{2k} \cdot n$ many) variables for which U is a pivot tuple with respect to $\xi|_1^{2:a}$.

To conclude, we bound from below $\mathrm{Ind}(\xi^U) - \mathrm{Ind}(\xi)$ for every $U \in \mathcal{S}$. We set $\Delta \overset{\text{def}}{=} \mathrm{Ind}(\xi^U) - \mathrm{Ind}(\xi)$ and analyze it:

$$\Delta = \frac{1}{n} \sum_{i \in [n]} \left(\mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{v}) \sim \xi^U|_{2,3}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi^U|_1^{2,3:(\boldsymbol{a},\boldsymbol{v})}} [\boldsymbol{x}_i = 1]^2 \right] - \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}} [\boldsymbol{x}_i = 1]^2 \right] \right)$$

$$= \frac{1}{n} \sum_{i \in [n]} \left(\mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\mathop{\mathbf{E}}_{\boldsymbol{v} \sim \xi^U|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi^U|_1^{2,3:(\boldsymbol{a},\boldsymbol{v})}} [\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}} [\boldsymbol{x}_i = 1]^2 \right] \right] \right)$$

$$= \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\frac{1}{n} \sum_{i \in [n]} \left(\mathop{\mathbf{E}}_{\boldsymbol{v} \sim \xi^U|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi^U|_1^{2,3:(\boldsymbol{a},\boldsymbol{v})}} [\boldsymbol{x}_i = 1]^2 \right] - \mathop{\mathbf{E}}_{\boldsymbol{v} \sim \xi^U|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2,3:(\boldsymbol{a},\boldsymbol{v})}} [\boldsymbol{x}_i = 1] \right]^2 \right) \right]$$

$$\geq \mathop{\mathbf{Pr}}_{\boldsymbol{a} \sim \xi|_2} [K_U] \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2^{2:K_U}} \left[\frac{1}{n} \sum_{i \in I_{U,a}} \left(\mathop{\mathbf{E}}_{\boldsymbol{v} \sim \xi^U|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi^U|_1^{2,3:(\boldsymbol{a},\boldsymbol{v})}} [\boldsymbol{x}_i = 1]^2 \right] - \mathop{\mathbf{E}}_{\boldsymbol{v} \sim \xi^U|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2,3:(\boldsymbol{a},\boldsymbol{v})}} [\boldsymbol{x}_i = 1] \right]^2 \right) \right]$$

$$\geq \frac{\epsilon^2}{4k} \cdot \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2^{2:K_U}} \left[\frac{|I_{U,\boldsymbol{a}}|}{n} \cdot \left(\frac{\epsilon}{2k} \right)^2 \frac{\epsilon}{2k} \right] \geq \frac{\epsilon^6}{64k^5},$$

20

where in the last line we used Observation 2.35 for the variable $\mathbf{X}_{a,i} = \mathbf{Pr}_{\boldsymbol{x} \sim \xi|_1^{2,3:(a,v)}}[\boldsymbol{x}_i = 1]$ and the probability space $\boldsymbol{v} \sim \xi^U|_3^{2:a}$, for every $a \in K_U$ and $i \in I_{U,a}$, with $\alpha = \beta = \frac{\epsilon}{2k}$. ∎

3.5 From variable weak robustness to general robustness

In this section we prove that weak variable robustness implies even the seemingly much stronger notion of general robustness, as stated in the following lemma.

Lemma 3.32. *Fix $\gamma \in (0,1)$, $\ell \in \mathbb{N}$ and let ξ be a detailing of μ with respect to A. If ξ is $(\gamma/2, C \cdot \frac{\ell^8}{\gamma^{38}})$-weakly robust for some absolute constant $C > 1$, then it is (γ, ℓ)-robust.*

The above lemma is a direct consequence of the following lemma (which will be proved here), which is a slightly strengthened restatement of Lemma 3.32 in a "contrary form".

Lemma 3.33. *Fix $\gamma \in (0,1)$, let ξ be a detailing of μ with respect to A, and let ζ be a refinement of ξ with respect to $A \times B$ such that $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi) \geq \gamma$. Then, there exists $k = O\left(\frac{|B|^8}{\gamma^{38}}\right)$ such that the following holds. A uniformly random k-tuple $\boldsymbol{\alpha} \in [n]^k$ satisfies $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi^{\boldsymbol{\alpha}}) < \gamma/2$, with probability at least $1/2$.*

To prove Lemma 3.33, we need the following *progress lemma* to be proved in Section 3.5.1.

Lemma 3.34. *If ξ is a detailing of μ with respect to A, and ζ is a refinement of ξ with respect to $A \times B$ satisfying $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi) \geq \gamma$, then for at least $\Omega\left(\frac{\gamma^{19}n}{|B|^4}\right)$ many $i \in [n]$ we have $\mathrm{Ind}(\xi^{\{i\}}) - \mathrm{Ind}(\xi) \geq \Omega\left(\frac{\gamma^{19}}{|B|^4}\right)$.*

We now define some notations. Let $\boldsymbol{\alpha} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) \in [n]^k$ be a uniformly random k-tuple and define $\mathbf{X}_1, \ldots, \mathbf{X}_k$ to be the following random variables for every $t \in [k]$ (which depend on $\boldsymbol{\alpha}$),

$$\mathbf{X}_t = \begin{cases} 0, & \text{if } \mathrm{Ind}(\xi^{\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_t\}}) - \mathrm{Ind}(\xi^{\{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{t-1}\}}) < \rho \\ 1, & \text{otherwise} \end{cases},$$

where $\rho = \Omega\left(\frac{\gamma^{19}}{|B|^4}\right)$, with the implicit coefficient in this expression to be later determined in the proof of Lemma 3.34. Next, for every $t \in [k]$ we define the random variable $\mathbf{R}_t \in [n] \cup \{0\}$ as

$$\mathbf{R}_t = \begin{cases} \boldsymbol{x}_t, & \text{if } \mathbf{X}_t = 1 \\ 0, & \text{otherwise} \end{cases},$$

Now we have the following lemma which lower bounds the probability of the event $\mathbf{R}_t \neq 0$ conditioned on the values of $\mathbf{R}_1, \ldots, \mathbf{R}_{t-1}$ for every $t \in [k]$.

Lemma 3.35. *Let $t \in [k]$ and $r_1, \ldots, r_{t-1} \in [n] \cup \{0\}$. If $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi^{\{r_1,\ldots,r_{t-1}\}\setminus\{0\}}) \geq \gamma/2$, then*

$$\mathbf{Pr}_{\boldsymbol{\alpha} \sim [n]^k}[\mathbf{R}_t \neq 0 \mid \mathbf{R}_1 = r_1, \ldots, \mathbf{R}_{t-1} = r_{t-1}] \geq \rho,$$

where $\rho = \Omega\left(\frac{\gamma^{19}}{|B|^4}\right)$.

Proof: We apply Lemma 3.34 with the detailing $\xi^{\{r_1,\dots,r_{t-1}\}\setminus\{0\}}$, and $\gamma/2$ instead of γ. By our assumptions we have $\mathrm{Ind}\,(\zeta) - \mathrm{Ind}(\xi^{\{r_1,\dots,r_{t-1}\}\setminus\{0\}}) \geq \gamma/2$, so there exists a set $S \subseteq [n]$ of size at least $\Omega\left(\frac{\gamma^{19}n}{|B|^4}\right)$, such that for any $x_t \in S$, we have

$$\mathrm{Ind}(\xi^{(\{r_1,\dots,r_{t-1}\}\setminus\{0\})\cup\{x_t\}}) - \mathrm{Ind}(\xi^{\{r_1,\dots,r_{t-1}\}\setminus\{0\}}) \geq \Omega\left(\frac{\gamma^{19}}{|B|^4}\right). \tag{1}$$

Note that if $\boldsymbol{x}_t \in S$, then since Equation (1) holds, we must have that $\mathbf{X}_t = 1$ and $\mathbf{R}_t = \boldsymbol{x}_t \neq 0$. As the probability that $\boldsymbol{x}_t \in S$ is at least $\Omega\left(\frac{\gamma^{19}}{|B|^4}\right)$ the lemma follows. ∎

Now we will use the following lemma from [AFL24].

Lemma 3.36 ([AFL24]). *Let $\mathcal{G} \subset \mathbb{R}^*$ be a set of goal sequences, satisfying that if u is a prefix of v and $u \in \mathcal{G}$ then $v \in \mathcal{G}$. Additionally, let $\mathbf{R}_1, \dots, \mathbf{R}_M$ be a set of random variables and $p_1, \dots, p_M$ be values in $[0,1]$, such that for every $1 \leq t \leq M$ and $v = (r_1, \dots, r_{t-1}) \in \mathbb{R}^{t-1} \setminus \mathcal{G}$ (that can happen with positive probability), we have $\mathbf{Pr}\left[\mathbf{R}_i \neq 0 \mid \mathbf{R}_1 = r_1, \dots, \mathbf{R}_{t-1} = r_{t-1}\right] \geq p_t$. For every $1 \leq t \leq M$, let $\mathbf{X}_t \in \{0,1\}$ be an indicator for $\mathbf{R}_t \neq 0$ and $\mathbf{X} = \sum_{t=1}^{M} \mathbf{X}_t$. Under these premises, for every $0 < \delta < 1$, the following holds:*

$$\mathbf{Pr}\left[((\mathbf{R}_1,\dots,\mathbf{R}_M) \notin \mathcal{G}) \wedge \left(\mathbf{X} < (1-\delta)\sum_{t=1}^{M} p_t\right)\right] < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\sum_{t=1}^{M} p_t}$$

Proof of Lemma 3.33: We define the set of goals as follows

$$\mathcal{G} \stackrel{\mathrm{def}}{=} \left\{(r_1, \dots, r_k) \in ([n] \cup \{0\})^k : \mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi^{\{r_1,\dots,r_k\}\setminus\{0\}}) < \gamma/2\right\}$$

Set $M = k$ for some k to be defined later. Let $p_1 = \dots = p_k = \rho$, where $\rho = \frac{\gamma^{19}}{C\cdot|B|^4}$ for some large enough absolute constant $C > 1$. By following Lemma 3.36 and Lemma 3.35, we have that for every $\delta \in (0,1)$,

$$\mathbf{Pr}\left[((\mathbf{R}_1,\dots,\mathbf{R}_k) \notin \mathcal{G}) \wedge \left(\mathbf{X} < (1-\delta)\sum_{t=1}^{k} \rho\right)\right] < \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\sum_{t=1}^{k} \rho}.$$

As $\left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right) \leq e^{-\frac{\delta^2}{2}}$ for $\delta \in (0,1)$, we have:

$$\mathbf{Pr}\left[((\mathbf{R}_1,\dots,\mathbf{R}_k) \notin \mathcal{G}) \wedge (\mathbf{X} < (1-\delta)\rho \cdot k]\right] < e^{-\frac{\delta^2 \rho k}{2}},$$

which implies that

$$\mathbf{Pr}\left[((\mathbf{R}_1,\dots,\mathbf{R}_k) \in \mathcal{G}) \vee (\mathbf{X} \geq (1-\delta)\rho \cdot k]\right] \geq 1 - e^{-\frac{\delta^2 \rho k}{2}}.$$

Next, we show that the event $\mathbf{X} \geq (1-\delta)\rho \cdot k$ is empty. We set $k \stackrel{\mathrm{def}}{=} \lceil 8/\rho^2 \rceil = O\left(\frac{|B|^8}{\gamma^{38}}\right)$ and $\delta = 1/2$. Then, $(1-\delta)\rho \cdot k \geq 1/\rho + 1$, and by the definition of the random variables $\mathbf{X}_t$ we have that

$$\mathrm{Ind}(\xi^{\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}\setminus\{0\}}) \geq (1-\delta)\rho \cdot k \cdot \rho > 1,$$

which is a contradiction to the fact that the index of a detailing is bounded by 1.

This implies that for the setting of $k = \lceil 8/\rho^2 \rceil$ and $\delta = 1/2$, $\mathbf{Pr}_{\boldsymbol{\alpha} \sim [n]^k}[(\mathbf{R}_1,\dots,\mathbf{R}_k) \in \mathcal{G}] \geq 1/2$. Therefore, by the definition of $\mathcal{G}$, we have that with probability at least $1/2$ for a uniformly random k-tuple $\boldsymbol{\alpha} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ it holds that $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi^{\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}\setminus\{0\}}) < \gamma/2$ as required. ∎

3.5.1 Proof of the progress lemma

To prove Lemma 3.34 we begin with the following.

Lemma 3.37. *Fix $\gamma \in (0,1)$ and suppose that there is a detailing ξ of μ with respect to A and a refinement ζ of ξ over $A \times B$ such that $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi) \geq \gamma$. Then the following hold:*

1. *There exists a set $A' \subseteq A$ such that $\xi|_2(A') \geq \gamma/2$, and for every $a \in A'$ there exists $b_a \in B$ (depending on a) such that $\frac{\zeta|_{2,3}(a,b_a)}{\xi|_2(a)} = \zeta|_3^{2:a}(b_a) \geq \frac{\gamma}{4|B|}$ for which*

$$
\mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(a,b_a)}} [\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}} [\boldsymbol{x}_i = 1]^2 \right] \geq \gamma/4.
$$

2. *For every fixed $a \in A'$ and $b \in B$, $\zeta|_3^{2:a}(b) \leq 1 - \gamma/8$.*

Proof: We start by proving the first item. By the premise that $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi) \geq \gamma$ we have:

$$
\gamma \leq \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \zeta|_{2,3}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}} [\boldsymbol{x}_i = 1]^2 \right] - \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}} [\boldsymbol{x}_i = 1]^2 \right] \right]
$$

$$
= \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\mathop{\mathbf{E}}_{\boldsymbol{b} \sim \zeta|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}} [\boldsymbol{x}_i = 1]^2 \right] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}} [\boldsymbol{x}_i = 1]^2 \right] \right]
$$

$$
= \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \xi|_2} \left[\mathop{\mathbf{E}}_{\boldsymbol{b} \sim \zeta|_3^{2:\boldsymbol{a}}} \left[\mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}} [\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}} [\boldsymbol{x}_i = 1]^2 \right] \right] \right].
$$

Therefore, by using the reverse Markov's inequality (Lemma 2.36) we have that there exists a set $A' \subseteq A$ with $\mathbf{Pr}_{\xi|_2}[A'] \geq \gamma/2$ such that for every $a \in A'$ it holds that

$$
\gamma/2 \leq \mathop{\mathbf{E}}_{\boldsymbol{b} \sim \zeta|_3^{2:a}} \left[\mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(a,\boldsymbol{b})}} [\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}} [\boldsymbol{x}_i = 1]^2 \right] \right]. \tag{2}
$$

By using the reverse Markov's inequality again, for every $a \in A'$ there exists a set B_a with $\mathbf{Pr}_{\zeta|_3^{2:a}}[B_a] \geq \gamma/4$ so that for every $b \in B_a$,

$$
\gamma/4 \leq \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(a,b)}} [\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}} [\boldsymbol{x}_i = 1]^2 \right]. \tag{3}
$$

Finally, by an averaging argument (noting that $|B_a| \leq |B|$), for every $a \in A'$ there exists $b_a \in B_a$ with $\zeta|_3^{2:a}(b_a) \geq \gamma/4|B|$ which satisfies Equation (3), completing the proof of the first item.

For the second item, note that from Equation (2) for any fixed $a \in A'$ we have that

$$
\gamma/2 \leq \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{\boldsymbol{b} \sim \zeta|_3^{2:a}} \left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \zeta|_1^{2,3:(a,\boldsymbol{b})}} [\boldsymbol{x}_i = 1]^2 \right] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}} [\boldsymbol{x}_i = 1]^2 \right] = \mathrm{Ind}(\zeta|^{2:a}) - \mathrm{Ind}(\xi|^{2:a}). \tag{4}
$$

Suppose that there is $b^* \in B$ such that $\zeta|_3^{2:a}(b^*) > 1 - \gamma/8$. Then,

$$\mathrm{Ind}(\zeta|^{2:a}) = \mathop{\mathbf{E}}_{i\sim[n]}\left[\mathop{\mathbf{E}}_{\boldsymbol{b}\sim\zeta|_3^{2:a}}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,\boldsymbol{b})}}[\boldsymbol{x}_i = 1]^2\right]\right] = \mathop{\mathbf{E}}_{\boldsymbol{b}\sim\zeta|_3^{2:a}}\left[\mathrm{Ind}(\zeta|^{2,3:(a,\boldsymbol{b})})\right]$$

$$= \zeta|_3^{2:a}(b^*)\cdot\mathrm{Ind}(\zeta|^{2,3:(a,b^*)}) + \sum_{b\neq b^*\in B}\zeta|_3^{2:a}(b)\cdot\mathrm{Ind}(\zeta|^{2,3:(a,b)})$$

$$\leq \mathrm{Ind}(\zeta|^{2,3:(a,b^*)}) + \frac{\gamma}{8}. \tag{5}$$

On the other hand,

$$\mathrm{Ind}(\xi|^{2:a}) = \mathop{\mathbf{E}}_{i\sim[n]}\left[\mathop{\mathbf{E}}_{\boldsymbol{b}\sim\zeta|_3^{2:a}}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,\boldsymbol{b})}}[\boldsymbol{x}_i = 1]\right]^2\right] \geq \mathop{\mathbf{E}}_{i\sim[n]}\left[\left(\zeta|_3^{2:a}(b^*)\cdot\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b^*)}}[\boldsymbol{x}_i = 1]\right)^2\right]$$

$$\geq (1-\gamma/8)\mathrm{Ind}(\zeta|^{2,3:(a,b^*)}).$$

Plugging in Equation (5) we get that

$$\mathrm{Ind}(\zeta|^{2:a}) \leq \frac{\mathrm{Ind}(\xi|^{2:a})}{(1-\gamma/8)} + \gamma/8 \leq \mathrm{Ind}(\xi|^{2:a})(1+\gamma/8) + \gamma/8 \leq \mathrm{Ind}(\xi|^{2:a}) + \frac{\gamma}{4},$$

which is a contradiction to Equation (4). ∎

Definition 3.38 (Complement outside b)**.** Given a detailing ζ of μ over $A \times B$, for every $a \in A$ and $b \in B$ we define $\kappa_{a,b}$, the *complement outside b*, as the distribution

$$\zeta|_1^{2:a,3:B\setminus\{b\}} = \frac{1}{1 - \zeta|_3^{2:a}(b)}\sum_{b'\neq b\in B}\zeta|_3^{2:a}(b')\cdot\zeta|_1^{2,3:(a,b')}.$$

Lemma 3.39. *Fix $(a, b) \in A \times B$ and $\gamma \in (0, 1)$. If*

$$\mathop{\mathbf{E}}_{i\sim[n]}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b)}}[\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\xi|_1^{2:a}}[\boldsymbol{x}_i = 1]^2\right] \geq \gamma,$$

then there exists a set $S_a \subseteq [n]$ of size at least $\gamma n/2$ such that for every $i \in S_a$ it holds that

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b)}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\kappa_{a,b}}[\boldsymbol{x}_i = 1] \geq \gamma/4.$$

Proof: By the reverse Markov inequality there exists a set $S_a \subseteq [n]$ of size at least $\gamma n/2$ such that for each $i \in S_a$ it holds that $\mathbf{Pr}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b)}}[\boldsymbol{x}_i = 1]^2 - \mathbf{Pr}_{\boldsymbol{x}\sim\xi|_1^{2:a}}[\boldsymbol{x}_i = 1]^2 \geq \gamma/2$. Since $c^2 - d^2 \leq 2|c-d|$ for $c, d \in [0, 1]$, for all $i \in S_a$ we have

$$\mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b)}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x}\sim\xi|_1^{2:a}}[\boldsymbol{x}_i = 1] \geq \gamma/4. \tag{6}$$

Note that since $\mathbf{Pr}_{\boldsymbol{x}\sim\xi|_1^{2:a}}[\boldsymbol{x}_i = 1] = \zeta|_3^{2:a}(b)\cdot\mathbf{Pr}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b)}}[\boldsymbol{x}_i = 1] + (1-\zeta|_3^{2:a}(b))\,\mathbf{Pr}_{\boldsymbol{x}\sim\kappa_{a,b}}[\boldsymbol{x}_i = 1]$ and $\mathbf{Pr}_{\boldsymbol{x}\sim\zeta|_1^{2,3:(a,b)}}[\boldsymbol{x}_i = 1] > \mathbf{Pr}_{\boldsymbol{x}\sim\xi|_1^{2:a}}[\boldsymbol{x}_i = 1]$, we must have that $\mathbf{Pr}_{\boldsymbol{x}\sim\kappa_{a,b}}[\boldsymbol{x}_i = 1] < \mathbf{Pr}_{\boldsymbol{x}\sim\xi|_1^{2:a}}[\boldsymbol{x}_i = 1]$. Combining that with Equation (6) implies the lemma. ∎

Lemma 3.40. *For any $a \in A$, if there exists $b_a \in B$ and $S_a \subseteq [n]$ such that for all $i \in S_a$ it holds that*

$$\Pr_{\boldsymbol{x} \sim \zeta|_1^{2,3:(a,b_a)}} [\boldsymbol{x}_i = 1] - \Pr_{\boldsymbol{x} \sim \kappa_{a,b_a}} [\boldsymbol{x}_i = 1] \geq \gamma,$$

then it holds that $\mathbf{Var}_{\boldsymbol{x} \sim \xi|_1^{2:a}} \left[\sum_{i \in S_a} \boldsymbol{x}_i \right] \geq \gamma^2 \delta_1 \delta_2 |S_a|^2$, where $\delta_1 = \zeta|_3^{2:a}(b_a)$ and $\delta_2 = 1 - \delta_1$.

Proof: Let $\mathbf{X} \stackrel{\text{def}}{=} \sum_{i \in S_a} \boldsymbol{x}_i$ and let $\mathbf{Z} \in \{1, 2\}$ be a random variable such that $\mathbf{Z} = 1$ with probability δ_1 and $\mathbf{Z} = 2$ with probability δ_2. We can think of the distribution of $\boldsymbol{x} \sim \xi|_1^{2:a}$ as first drawing $\mathbf{Z}$ and then drawing $\boldsymbol{x}$ from $\zeta|_1^{2,3:(a,b_a)}$ or κ_{a,b_a} based on whether $\mathbf{Z} = 1$ or $\mathbf{Z} = 2$, respectively. By using Cauchy-Schwartz inequality (Lemma 2.34),

$$\mathbf{E}_{\boldsymbol{x} \sim \xi|_1^{2:a}} [\mathbf{X}^2] = \mathbf{E}_{\mathbf{Z}} \left[\mathbf{E}_{\boldsymbol{x}} [\mathbf{X}^2 \mid \mathbf{Z}] \right] \geq \mathbf{E}_{\mathbf{Z}} \left[\mathbf{E}_{\boldsymbol{x}} [\mathbf{X} \mid \mathbf{Z}]^2 \right].$$

By the premise of the lemma we have that

$$\mathbf{E}_{\boldsymbol{x}} [\mathbf{X} \mid \mathbf{Z} = 1] - \mathbf{E}_{\boldsymbol{x}} [\mathbf{X} \mid \mathbf{Z} = 2] \geq \gamma |S_a|.$$

Combining the above (and using $\delta_1 + \delta_2 = 1$),

$$\begin{aligned}
\mathbf{Var}_{\boldsymbol{x} \sim \xi|_1^{2:a}} [\mathbf{X}] = \mathbf{E}[\mathbf{X}^2] - \mathbf{E}[\mathbf{X}]^2 &\geq \mathbf{E}_{\mathbf{Z}} \left[\mathbf{E}[\mathbf{X} \mid \mathbf{Z}]^2 \right] - \left(\mathbf{E}_{\mathbf{Z}} [\mathbf{E}[\mathbf{X} \mid \mathbf{Z}]] \right)^2 \\
&= \delta_1 \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 1]^2 + \delta_2 \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 2]^2 - \delta_1^2 \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 1]^2 - \delta_2^2 \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 2]^2 \\
&\quad - 2 \delta_1 \delta_2 \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 1] \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 2] \\
&= \delta_1 (1 - \delta_1) \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 1]^2 + \delta_2 (1 - \delta_2) \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 2]^2 - 2 \delta_1 \delta_2 \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 1] \cdot \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 2] \\
&= \delta_1 \delta_2 \left(\mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 1] - \mathbf{E}[\mathbf{X} \mid \mathbf{Z} = 2] \right)^2 \\
&\geq \gamma^2 \delta_1 \delta_2 |S_a|^2
\end{aligned}$$

This completes the proof of the lemma. $\blacksquare$

Now we prove that given that the conclusion of Lemma 3.40 holds, that is, $\mathbf{Var}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\sum_{i \in S_a} \boldsymbol{x}_i]$ is large, there exists a subset of indices $S_a' \subseteq S_a$ such that for every index $i \in S_a'$, there is a large number of indices $j \in S_a$ with large covariance between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$.

Lemma 3.41. *Let $S_a \subseteq [n]$ be such that $\mathbf{Var}_{\boldsymbol{x} \sim \xi|_1^{2:a}} \left[\sum_{i \in S_a} \boldsymbol{x}_i \right] \geq \beta |S_a|^2$ for some $\beta > 0$. Then there is a subset $S_a' \subseteq S_a$ with $|S_a'| \geq \beta |S_a|/4$, such that for every $i \in S_a'$ there exists a set $A_{i,a} \subseteq S_a$ with at least $\beta |S_a|/2$ many indices, where for all $j \in A_{i,a}$ we have $\mathbf{Cov}_{\boldsymbol{x} \sim \xi|_1^{2:a}} (\boldsymbol{x}_i, \boldsymbol{x}_j) \geq \beta/4$.*

Proof: For every $i \in S_a$ we define

$$A_{i,a} \stackrel{\text{def}}{=} \{ j \in S_a : \mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) > \beta/4 \},$$

and let

$$S_a' \stackrel{\text{def}}{=} \{ i \in S_a : |A_{i,a}| \geq \beta |S_a|/2 \}.$$

Then, for all $i, j \in S_a$, using the fact that $\mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq 1$ for all $i, j \in [n]$ we have:

$$\beta |S_a|^2 \leq \mathop{\mathbf{Var}}_{\boldsymbol{x} \sim \xi|_1^{2:a}} \left[\sum_{i \in S_a} \boldsymbol{x}_i \right] = \sum_{i,j \in S_a} \mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

$$= \sum_{i \in S_a'} \left(\sum_{j \in A_{i,a}} \mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{j \notin A_{i,a}} \mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) \right) + \sum_{i \notin S_a'} \left(\sum_{j \in A_{i,a}} \mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sum_{j \notin A_{i,a}} \mathbf{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_j) \right)$$

$$\leq |S_a'||S_a| + |S_a \setminus S_a'|(\beta |S_a|/2 + \beta |S_a|/4)$$

$$\leq |S_a'||S_a| + |S_a|^2 \cdot \frac{3\beta}{4},$$

This implies that $\frac{\beta}{4}|S_a| \leq |S_a'|$, which finishes the proof. ∎

Lemma 3.42. *Fix $\gamma \in (0,1)$ and suppose that there is a detailing ξ of μ with respect to A and a refinement ζ of ξ over $A \times B$ such that $\mathrm{Ind}(\zeta) - \mathrm{Ind}(\xi) \geq \gamma$. Then there exists a set A' with $\mathbf{Pr}_{\xi|_2}[A'] \geq \gamma/2$, and a set $S_a' \subseteq [n]$ of size at least $\frac{\gamma^5}{2^{10}|B|} n$ for every $a \in A'$, so that for all $i^* \in S_a'$ we have*

$$\mathrm{Ind}((\xi|^{2:a})^{\{i^*\}}) - \mathrm{Ind}(\xi|^{2:a}) \geq \frac{\gamma^{13}}{2^{28}|B|^3}$$

Proof: Before constructing A' and S_a', let us prove an index lower bound for every $a \in A$ and $i^* \in [n]$. Let $\mathbf{Z}_i^a$ be a random variable defined by the following process. We draw $\boldsymbol{j} \in \{0,1\}$ such that $\boldsymbol{j} = 1$ with probability $\mathbf{Pr}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*} = 1]$ and let $\mathbf{Z}_i^a = \mathbf{Pr}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1 \mid \boldsymbol{x}_{i^*} = \boldsymbol{j}]$. Note that $\mathbf{E}_{\boldsymbol{j}}[\mathbf{Z}_i^a] = \mathbf{Pr}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1]$ and therefore,

$$\mathop{\mathbf{Var}}_{\boldsymbol{j}}[\mathbf{Z}_i^a] = \mathbf{E}_{\boldsymbol{j}} \left[(\mathbf{Z}_i^a - \mathbf{E}[\mathbf{Z}_i^a])^2 \right] = \sum_{\boldsymbol{j} \in \{0,1\}} \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*} = \boldsymbol{j}] \left(\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1 \mid \boldsymbol{x}_{i^*} = \boldsymbol{j}] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1] \right)^2$$

$$\geq \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*} = 1] \left(\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1 \mid \boldsymbol{x}_{i^*} = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1] \right)^2$$

$$\geq \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*} = 1]^2 \left(\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1 \mid \boldsymbol{x}_{i^*} = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1] \right)^2$$

$$= \left(\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1 \wedge \boldsymbol{x}_{i^*} = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1] \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*} = 1] \right)^2$$

$$= \left(\mathop{\mathbf{E}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i \boldsymbol{x}_{i^*}] - \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i] \mathop{\mathbf{E}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*}] \right)^2$$

$$= \left(\mathbf{Cov}_{\boldsymbol{x} \sim \xi|_1^{2:a}}(\boldsymbol{x}_{i^*}, \boldsymbol{x}_i) \right)^2.$$

Thus, by the definition of the index of a distribution,

$$\mathrm{Ind}((\xi|^{2:a})^{\{i^*\}}) - \mathrm{Ind}(\xi|^{2:a}) = \mathop{\mathbf{E}}_{i \sim [n]} \left[\sum_{\boldsymbol{j} \in \{0,1\}} \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_{i^*} = \boldsymbol{j}] \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1 \mid \boldsymbol{x}_{i^*} = \boldsymbol{j}]^2 - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:a}}[\boldsymbol{x}_i = 1]^2 \right]$$

$$= \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathbf{E}_{\boldsymbol{j}}[(\mathbf{Z}_i^a)^2] - \mathbf{E}_{\boldsymbol{j}}[\mathbf{Z}_i^a]^2 \right] = \mathop{\mathbf{E}}_{i \sim [n]} [\mathbf{Var}_{\boldsymbol{j}}[\mathbf{Z}_i^a]].$$

Now we go back to constructing out sets. We use Lemma 3.37 to obtain A', noting also b_a for every $a \in A'$. Then we use Lemma 3.39 for every $a \in A'$ along with b_a with parameter $\gamma/4$, noting the resulting set S_a, whose size is at least $\gamma n/8$. Next we use Lemma 3.40 over this with parameter $\gamma/16$, also recalling that in the notation of that lemma $\delta_1 \geq \gamma/4|B|$ and $\delta_2 \geq \gamma/8$ (as guaranteed by Lemma 3.37) to obtain the bound $\mathbf{Var}_{\boldsymbol{x} \sim \xi|_1^{2:a}} \left[\sum_{i \in S_a} \boldsymbol{x}_i \right] \geq \gamma^2 \delta_1 \delta_2 |S_a|^2 \geq \frac{\gamma^4}{32|B|} |S_a|^2$.

Now we use Lemma 3.41, with $\beta = \frac{\gamma^4}{32|B|}$ to obtain the set S'_a and the sets $A_{i,a}$ for all $i \in S'_a$. Note that the obtained bounds here are $|S'_a| \geq \frac{\gamma^4}{32|B|}|S_a| \geq \frac{\gamma^5}{2^{10}|B|}n$ and $|A_{i,a}| \geq \frac{\gamma^5}{2^{10}|B|}n$, where for $i \in S'_a$ and $j \in A_{i,a}$ we have $\mathbf{Cov}_{\boldsymbol{x} \sim \xi|_1^{2:a}} (\boldsymbol{x}_i, \boldsymbol{x}_j) \geq \frac{\gamma^4}{2^9|B|}$. Thus, for every $a \in A'$ and $i^* \in S'_a$ we have

$$
\mathrm{Ind}((\xi|^{2:a})^{\{i^*\}}) - \mathrm{Ind}(\xi|^{2:a}) = \frac{1}{n} \sum_{i \in [n]} \mathbf{Var}_j[\mathbf{Z}_i^a] \geq \frac{1}{n} \sum_{i \in [n]} \left(\mathbf{Cov}_{\boldsymbol{x} \sim \xi|_1^a}(\boldsymbol{x}_{i^*}, \boldsymbol{x}_i) \right)^2
$$

$$
\geq \frac{1}{n} \sum_{i \in A_{i^*,a}} \left(\frac{\gamma^4}{2^9|B|} \right)^2 = \frac{\gamma^{13}}{2^{28}|B|^3},
$$

completing our proof. ∎

Proof of Lemma 3.34: We start by invoking Lemma 3.42 to obtain the sets A' and S'_a for every $a \in A$ that are guaranteed by it. Now note that

$$
\mathbf{E}_{\boldsymbol{i} \sim [n]} \left[\mathrm{Ind}(\xi^{\{i\}}) - \mathrm{Ind}(\xi) \right] = \mathbf{E}_{\boldsymbol{i} \sim [n]} \left[\mathbf{E}_{\boldsymbol{a} \sim \xi|_2} \left[\mathrm{Ind}(\xi^{\{i\}}) - \mathrm{Ind}(\xi) \right] \right]
$$

$$
= \mathbf{E}_{\boldsymbol{a} \sim \xi|_2} \left[\mathbf{E}_{\boldsymbol{i} \sim [n]} \left[\mathrm{Ind}(\xi^{\{i\}}) - \mathrm{Ind}(\xi) \right] \right]
$$

$$
\geq \mathbf{Pr}_{\xi|_2}[A'] \mathbf{E}_{\boldsymbol{a} \sim \xi|_2^{2:A'}} \left[\frac{|S'_a|}{n} \mathbf{E}_{\boldsymbol{i} \sim S'_a} \left[\mathrm{Ind}(\xi^{\{i\}}) - \mathrm{Ind}(\xi) \right] \right]
$$

$$
\geq \frac{\gamma}{2} \mathbf{E}_{\boldsymbol{a} \sim \xi|_2^{2:A'}} \left[\frac{\gamma^5}{2^{10}|B|} \mathbf{E}_{\boldsymbol{i} \sim S'_a} \left[\frac{\gamma^{13}}{2^{28}|B^3|} \right] \right] = \frac{\gamma^{19}}{2^{39}|B|^4}.
$$

To conclude, we use the reverse Markov inequality to obtain that there exists a set of at least $\frac{\gamma^{19}}{2^{40}|B|^4}n$ indices i for which $\mathrm{Ind}(\xi^{\{i\}}) - \mathrm{Ind}(\xi) \geq \frac{\gamma^{19}}{2^{40}|B|^4}$ holds. ∎

4 Predicting the acceptance probability of a canonical tester

In this section we show that knowing the statistics of a good detailing is sufficient to approximate the acceptance probability of a property testing algorithm without actually running it.

4.1 Simulation of a canonical tester

We describe here a randomized process, `Simulate` (see Figure 1), that given the statistic of a detailing ξ of a distribution μ provides a simulated result of a canonical query pattern, without making any further queries to μ (eventually we will only use queries to find ξ and its statistic).

From now on we denote by $\mathcal{D}_{\mathsf{sim}}(\eta, \Lambda)$ the distribution over the output of `Simulate`(s, q, η, Λ) (note that this output is a randomized $s \times q$ matrix). We now show that given a robust enough detailing

---

Procedure `Simulate`(s, q, η, Λ)

Input: Integers $s, q \in \mathbb{N}$ corresponding to samples and queries, weight distribution η and type distribution Λ of a detailing ξ.
Output: A matrix $\mathbf{M} \in \{0, 1\}^{s \times q}$.

1. For each $j \in [q]$, draw $\boldsymbol{t}_j \sim \Lambda$ independently.

2. For each $i \in [s]$, draw $\boldsymbol{a}_i \sim \eta$ independently.

3. For every $(i, j) \in [s] \times [q]$, draw $\mathbf{M}_{i,j} \sim \mathrm{Ber}(\boldsymbol{t}_j(\boldsymbol{a}_i))$ independently.

4. **Return M**

Figure 1: Description of the `Simulate` procedure.

of μ, this simulated distribution is indeed close to the distribution $\mathcal{D}_{\mathsf{test}}$ generated by an actual canonical $s \times q$ query pattern.

Lemma 4.1. *Fix $s, q \in \mathbb{N}$ and $\epsilon \in (0, 1)$. If ξ is $\left(\frac{\epsilon}{3(s+1)}, q\right)$-good and $n \geq \frac{6q^2(s+1)}{\epsilon}$, then $d_{\mathrm{TV}}(\mathcal{D}_{\mathsf{sim}}(\eta, \Lambda), \mathcal{D}_{\mathsf{test}}) \leq \epsilon$.*

Proof: Let $\epsilon' = \frac{\epsilon}{3 \cdot (s+1)}$ and consider the following coupling $\mathbf{T}^*$ (see Observation 2.10) between $\mathcal{D}_{\mathsf{sim}}(\eta, \Lambda)$ and $\mathcal{D}_{\mathsf{test}}$:

We first draw $\boldsymbol{j}_1, \ldots, \boldsymbol{j}_q$ uniformly and independently from $[n]$ (with repetitions), and let $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_q$ be their respective types in ξ. Note that $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_q$ are distributed the same as independent draws from Λ. Then, we draw $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s$ independently from $\xi|_2$. Also, we let $\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q$ be equal to $\boldsymbol{j}_1, \ldots, \boldsymbol{j}_q$ if they contain no repetition, and otherwise let $(\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q)$ be uniformly drawn *without repetitions* from $[n]^q$ (independently of $\boldsymbol{j}_1, \ldots, \boldsymbol{j}_q$).

For every $i \in [s]$, let $\mathbf{T}_i$ denote an optimal transfer distribution between $\prod_{j \in [q]} \mathrm{Ber}(\boldsymbol{t}_j(\boldsymbol{a}_i))$ and $\xi|^{2:\boldsymbol{a}_i}_{\{\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q\}}$. That is, $\mathbf{Pr}_{(\boldsymbol{x}, \tilde{\boldsymbol{x}}) \sim \mathbf{T}_i}[\boldsymbol{x} \neq \tilde{\boldsymbol{x}}] = d_{\mathrm{TV}}(\prod_{j \in [q]} \mathrm{Ber}(\boldsymbol{t}_j(\boldsymbol{a}_i)), \xi|^{2:\boldsymbol{a}_i}_{\{\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q\}})$. To obtain $\mathbf{M}$ and $\widetilde{\mathbf{M}}$, for every $i \in [s]$ independently, we draw $((\mathbf{M}_{i,1}, \ldots, \mathbf{M}_{i,q}), (\widetilde{\mathbf{M}}_{i,1}, \ldots, \widetilde{\mathbf{M}}_{i,q}))$ from $\mathbf{T}_i$.

In addition, we define the following random events:

- $\boldsymbol{\mathcal{E}}_0$ is the event that $(\boldsymbol{j}_1, \ldots, \boldsymbol{j}_q) = (\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q)$.

- $\boldsymbol{\mathcal{E}}_1$ is the event that $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s \in J$ where J is the good set as per Definition 3.25.

- $\boldsymbol{\mathcal{E}}_2$ is the event that the q-tuple $(\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q)$ is ϵ'-independent with respect to all $\xi_{\boldsymbol{a}_1}, \ldots, \xi_{\boldsymbol{a}_s}$.

- $\boldsymbol{\mathcal{E}}_3$ is the event that $(\mathbf{M}_{i,1}, \ldots, \mathbf{M}_{i,q}) = (\widetilde{\mathbf{M}}_{i,1}, \ldots, \widetilde{\mathbf{M}}_{i,q})$ for all $i \in [s]$.

We now note that

$$
\begin{aligned}
d_{\mathrm{TV}}(\mathcal{D}_{\mathsf{sim}}, \mathcal{D}_{\mathsf{test}}) &\leq \mathbf{Pr}_{\mathbf{T}^*}[\mathbf{M} \neq \widetilde{\mathbf{M}}] = \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_3] \\
&\leq \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_0] + \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_1] + \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_2 \wedge \boldsymbol{\mathcal{E}}_1] + \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_3 \wedge \boldsymbol{\mathcal{E}}_0 \wedge \boldsymbol{\mathcal{E}}_1 \wedge \boldsymbol{\mathcal{E}}_2] \\
&\leq \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_0] + \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_1] + \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_2 \mid \boldsymbol{\mathcal{E}}_1] + \mathbf{Pr}[\neg \boldsymbol{\mathcal{E}}_3 \mid \boldsymbol{\mathcal{E}}_0 \wedge \boldsymbol{\mathcal{E}}_1 \wedge \boldsymbol{\mathcal{E}}_2]
\end{aligned}
$$

28

By the fact that the detailing is (ϵ', q)-good, using a union bound we get $\mathbf{Pr}[\neg\boldsymbol{\mathcal{E}}_1] \leq s \cdot \epsilon'$. Similarly, the probability of a q-tuple $(\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q)$ to be ϵ'-independent with respect to $\xi|^{2:\boldsymbol{a}_i}$ is at least $1 - \epsilon'$, and thus the probability that the q-tuple is ϵ'-independent with respect to all $\xi|^{2:\boldsymbol{a}_1}, \ldots, \xi|^{2:\boldsymbol{a}_s}$ is greater than $1 - s \cdot \epsilon'$ (equivalently $\mathbf{Pr}[\neg\boldsymbol{\mathcal{E}}_2|\boldsymbol{\mathcal{E}}_1] \leq s \cdot \epsilon'$). To bound the probability of $\neg\boldsymbol{\mathcal{E}}_0$, note that if $n \geq 2q^2/\epsilon'$, then by the birthday paradox $\mathbf{Pr}[\neg\boldsymbol{\mathcal{E}}_0] \leq \frac{q(q-1)}{2n} \leq \epsilon'$. It is left to bound $\mathbf{Pr}[\neg\boldsymbol{\mathcal{E}}_3|\boldsymbol{\mathcal{E}}_0 \wedge \boldsymbol{\mathcal{E}}_1 \wedge \boldsymbol{\mathcal{E}}_2]$.

$$\mathbf{Pr}[\neg\boldsymbol{\mathcal{E}}_3|\boldsymbol{\mathcal{E}}_0 \wedge \boldsymbol{\mathcal{E}}_1 \wedge \boldsymbol{\mathcal{E}}_2] \leq \sum_{i \in [s]} \Pr_{(\mathbf{M}_{i,*}, \widetilde{\mathbf{M}}_{i,*}) \sim \mathbf{T}_i} \left[\mathbf{M}_{i,*} \neq \widetilde{\mathbf{M}}_{i,*} \mid \boldsymbol{\mathcal{E}}_0 \wedge \boldsymbol{\mathcal{E}}_1 \wedge \boldsymbol{\mathcal{E}}_2 \right]$$

$$\leq \sum_{i \in [s]} d_{\mathrm{TV}} \left(\prod_{j \in [q]} \mathrm{Ber}(\boldsymbol{t}_j(\boldsymbol{a}_i)), \xi|^{2:\boldsymbol{a}_i}_{\{\boldsymbol{j}'_1, \ldots, \boldsymbol{j}'_q\}} \right) \leq s \cdot \epsilon',$$

where the last inequality follows from the q-tuple being ϵ'-independent with respect to all $\xi_{\boldsymbol{a}_i}$ with $i \in [s]$. Overall, we get

$$d_{\mathrm{TV}}(\mathcal{D}_{\mathsf{sim}}(\eta, \Lambda), \mathcal{D}_{\mathsf{test}}) \leq s \cdot \epsilon' + 2s \cdot \epsilon' + \epsilon' = (3s+1)\epsilon' \leq \epsilon.$$

∎

However, our eventual estimation algorithm will not be able to accurately find η and Λ, but only approximations thereof. For this reason we will need "continuity lemmas" for these quantities, showing that the output distribution $\mathcal{D}_{\mathsf{sim}}$ of $\mathtt{Simulate}$ will not degrade by much.

The next lemma handles the setting where we only have an approximation $\tilde{\Lambda}$ of the type distribution Λ, that is close enough to it in the η-weighted ℓ_1 Earth Mover distance.

Lemma 4.2. *For $q, s \in \mathbb{N}$ and $\epsilon \in (0, 1)$, if $d^\eta_{\mathrm{EM}}(\Lambda, \tilde{\Lambda}) \leq \frac{\epsilon}{s \cdot q}$ then $d_{\mathrm{TV}}(\mathcal{D}_{\mathsf{sim}}(\eta, \Lambda), \mathcal{D}_{\mathsf{sim}}(\eta, \tilde{\Lambda})) \leq \epsilon$.*

Proof: By the premise of the lemma we have that $d^\eta_{\mathrm{EM}}(\Lambda, \tilde{\Lambda}) \leq \frac{\epsilon}{s \cdot q}$, so let $\mathbf{T}$ be a the transfer distribution over $([0,1]^A)^2$ exhibiting the distance. We consider the following coupling $\mathbf{T}^*$ between $\mathcal{D}_{\mathsf{sim}}(\eta, \Lambda)$ and $\mathcal{D}_{\mathsf{sim}}(\eta, \tilde{\Lambda})$. Sample $(\boldsymbol{t}_1, \tilde{\boldsymbol{t}}_1), \ldots, (\boldsymbol{t}_q, \tilde{\boldsymbol{t}}_q) \sim \mathbf{T}$ and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s \sim \eta$, all independently. To set $\mathbf{M}$ and $\widetilde{\mathbf{M}}$, perform the following for every $i \in [s]$ and $j \in [q]$ independently:

(i) If $\boldsymbol{t}_j(\boldsymbol{a}_i) \leq \tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i)$, then we do the following:

    (a) with probability $\boldsymbol{t}_j(\boldsymbol{a}_i)$ set $(\mathbf{M}_{i,j}, \widetilde{\mathbf{M}}_{i,j}) = (1, 1)$,

    (b) with probability $\tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i) - \boldsymbol{t}_j(\boldsymbol{a}_i)$ set $(\mathbf{M}_{i,j}, \widetilde{\mathbf{M}}_{i,j}) = (0, 1)$,

    (c) with probability $1 - \tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i)$ set $(\mathbf{M}_{i,j}, \widetilde{\mathbf{M}}_{i,j}) = (0, 0)$.

(ii) If $\tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i) \leq \boldsymbol{t}_j(\boldsymbol{a}_i)$, then we do the following:

    (a) with probability $\tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i)$ set $(\mathbf{M}_{i,j}, \widetilde{\mathbf{M}}_{i,j}) = (1, 1)$,

    (b) with probability $\boldsymbol{t}_j(\boldsymbol{a}_i) - \tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i)$ set $(\mathbf{M}_{i,j}, \widetilde{\mathbf{M}}_{i,j}) = (1, 0)$,

    (c) with probability $1 - \boldsymbol{t}_j(\boldsymbol{a}_i)$ set $(\mathbf{M}_{i,j}, \widetilde{\mathbf{M}}_{i,j}) = (0, 0)$.

Using the above coupling, we have

$$d_{\text{TV}}(\mathcal{D}_{\text{sim}}(\eta, \Lambda), \mathcal{D}_{\text{sim}}(\eta, \tilde{\Lambda})) \leq \Pr_{(\mathbf{M}, \widetilde{\mathbf{M}}) \sim \mathbf{T}^*} [\mathbf{M} \neq \widetilde{\mathbf{M}}] \leq \sum_{(i,j) \in [s] \times [q]} \Pr_{(\mathbf{M}, \widetilde{\mathbf{M}}) \sim \mathbf{T}^*} [\mathbf{M}_{i,j} \neq \widetilde{\mathbf{M}}_{i,j}]$$

$$\leq \sum_{(i,j) \in [s] \times [q]} \mathop{\mathbf{E}}_{(\boldsymbol{t}_j, \tilde{\boldsymbol{t}}_j) \sim \mathbf{T}} \mathop{\mathbf{E}}_{\boldsymbol{a}_i \sim \eta} [|\boldsymbol{t}_j(\boldsymbol{a}_i) - \tilde{\boldsymbol{t}}_j(\boldsymbol{a}_i)|]$$

$$= \sum_{(i,j) \in [s] \times [q]} d_{\text{EM}}^{\eta}(\Lambda, \tilde{\Lambda}) \leq \epsilon.$$

∎

Next, we handle the setting where we only have a distribution $\tilde{\eta}$ that is close to η in the variation distance.

Lemma 4.3. *For any* $q, s \in \mathbb{N}$, $\epsilon \in (0,1)$, *if* $d_{\text{TV}}(\eta, \tilde{\eta}) \leq \frac{\epsilon}{s}$, *then* $d_{\text{TV}}(\mathcal{D}_{\text{sim}}(\eta, \Lambda), \mathcal{D}_{\text{sim}}(\tilde{\eta}, \Lambda)) \leq \epsilon$.

Proof: Let $\mathbf{T}$ be an optimal coupling (transfer distribution) between η and $\tilde{\eta}$. Namely, $d_{\text{TV}}(\eta, \tilde{\eta}) = \Pr_{(\boldsymbol{a}, \tilde{\boldsymbol{a}}) \sim \mathbf{T}}[\boldsymbol{a} \neq \tilde{\boldsymbol{a}}]$. We construct a coupling $\mathbf{T}^*$ between $\mathcal{D}_{\text{sim}}(\eta, \Lambda)$ and $\mathcal{D}_{\text{sim}}(\tilde{\eta}, \Lambda)$. First, draw $(\boldsymbol{a}_1, \tilde{\boldsymbol{a}}_1), \ldots, (\boldsymbol{a}_s, \tilde{\boldsymbol{a}}_s) \sim \mathbf{T}$ and $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_q \sim \Lambda$, all independently. We define $\mathbf{M}$ and $\widetilde{\mathbf{M}}$ as follows: For every $(i,j) \in [s] \times [q]$, if $\boldsymbol{a}_i = \tilde{\boldsymbol{a}}_i$, then draw a bit $\boldsymbol{b} \sim \text{Ber}(\boldsymbol{t}_j(\boldsymbol{a}_i))$ and set $\mathbf{M}_{i,j} = \widetilde{\mathbf{M}}_{i,j} = \boldsymbol{b}$. If $\boldsymbol{a}_i \neq \tilde{\boldsymbol{a}}_i$, then independently set $\mathbf{M}_{i,j} \sim \text{Ber}(\boldsymbol{t}_j(\boldsymbol{a}_i))$ and $\widetilde{\mathbf{M}}_{i,j} \sim \text{Ber}(\boldsymbol{t}_j(\tilde{\boldsymbol{a}}_i))$. Then,

$$d_{\text{TV}}(\mathcal{D}_{\text{sim}}(\eta, \Lambda), \mathcal{D}_{\text{sim}}(\tilde{\eta}, \Lambda)) \leq \Pr_{(\mathbf{M}, \widetilde{\mathbf{M}}) \sim \mathbf{T}^*} [\mathbf{M} \neq \widetilde{\mathbf{M}}] \leq \sum_{i \in [s]} \Pr_{(\boldsymbol{a}_i, \tilde{\boldsymbol{a}}_i) \sim \mathbf{T}} [\boldsymbol{a}_i \neq \tilde{\boldsymbol{a}}_i] \leq \epsilon.$$

∎

The following lemma bundles together the approximation results in this section.

Lemma 4.4. *Fix* $s, q \in \mathbb{N}$ *and* $\epsilon \in (0,1)$ *and suppose that* $d_{\text{EM}}^{\eta}(\Lambda, \tilde{\Lambda}) \leq \epsilon/3sq$ *and* $d_{\text{TV}}(\eta, \tilde{\eta}) \leq \epsilon/3s$. *If* $n \geq 18q^2(s+1)/\epsilon$ *and* ξ *is* $\left(\frac{\epsilon}{9(s+1)}, q\right)$-*good then* $d_{\text{TV}}(\mathcal{D}_{\text{sim}}(\tilde{\eta}, \tilde{\Lambda}), \mathcal{D}_{\text{test}}) \leq \epsilon$.

Proof: First note that by the triangle inequality

$$d_{\text{TV}}\left(\mathcal{D}_{\text{sim}}(\tilde{\eta}, \tilde{\Lambda}), \mathcal{D}_{\text{test}}\right) \leq d_{\text{TV}}\left(\mathcal{D}_{\text{sim}}(\tilde{\eta}, \tilde{\Lambda}), \mathcal{D}_{\text{sim}}(\eta, \tilde{\Lambda})\right) + d_{\text{TV}}\left(\mathcal{D}_{\text{sim}}(\eta, \tilde{\Lambda}), \mathcal{D}_{\text{sim}}(\eta, \Lambda)\right)$$
$$+ d_{\text{TV}}\left(\mathcal{D}_{\text{sim}}(\eta, \Lambda), \mathcal{D}_{\text{test}}\right).$$

Applying Lemma 4.3, Lemma 4.2 and Lemma 4.1, all with $\epsilon/3$ instead of ϵ, gives that every term in the above sum is bounded by $\epsilon/3$. ∎

4.2 Acceptance probability computation

With Lemma 4.4 in mind, we construct an "acceptance predictor" procedure, Accept-Probability (see Figure 2), which is intended to *calculate* the estimated acceptance probability of a test. It works by calculating (instead of actually running) the output distribution of Simulate. Recall that we denote the probability that the canonical tester with proximity parameter ϵ accepts μ with $\text{acc}_\epsilon(\mu)$.

Procedure Accept-Probability$(s, q, \eta, \Lambda, \epsilon')$

Input: Integers $s, q \in \mathbb{N}$ corresponding to samples and queries, weight distribution η, type distribution Λ of a detailing ξ of μ with respect to A and test proximity parameter $\epsilon' \in (0, 1)$.
Output: An estimated acceptance probability $\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda) \in [0, 1]$.

1. Let $\alpha_{\epsilon'} : \{0, 1\}^{s \times q} \to [0, 1]$ be the acceptance probability function of the canonical tester with proximity parameter ϵ', as per Definition 2.20.

2. For each $M \in \{0, 1\}^{s \times q}$, $a \in A^s$ and $t \in ([0, 1]^A)^q$ in the support of Λ compute

$$\widetilde{\mathcal{D}}(M, a, t) = \prod_{i \in [s]} \eta(a_i) \cdot \prod_{j \in [q]} \Lambda(t_j) \cdot \prod_{i,j \in [s] \times [q]} (\mathbf{1}_{M_{i,j}=1} \cdot t_j(a_i) + \mathbf{1}_{M_{i,j}=0} \cdot (1 - t_j(a_i)))$$

3. Set

$$\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda) = \sum_{(M,a,t) \in \{0,1\}^{s \times q} \times A^s \times [0,1]^{A \times q}} \widetilde{\mathcal{D}}(M, a, t) \cdot \alpha_{\epsilon'}(M) = \mathop{\mathbf{E}}_{\mathbf{M} \sim \widetilde{\mathcal{D}}|_1} [\alpha_{\epsilon'}(\mathbf{M})]$$

4. **Return** $\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda)$.

Figure 2: Description of the Accept-Probability procedure.

Lemma 4.5. *Fix $\epsilon' \in (0, 1)$ and let $\mathcal{P}$ be an index-invariant property that admits a canonical tester with proximity parameter ϵ' using sample complexity $s = s(\epsilon')$ and query complexity $q = q(\epsilon')$. Fix $\epsilon \in (0, 1)$, and let (η, Λ) be the parameters of an $\left(\frac{\epsilon}{9(s+1)}, q\right)$-good detailing of μ with respect to $U \subseteq [n]$. If $n \geq 18q^2(s+1)/\epsilon$, then given $(\widetilde{\eta}, \widetilde{\Lambda})$ such that $d_{\mathrm{TV}}(\eta, \widetilde{\eta}) \leq \frac{\epsilon}{3s}$ and $d^{\eta}_{\mathrm{EM}}(\Lambda, \widetilde{\Lambda}) \leq \frac{\epsilon}{3sq}$, the subroutine Accept-Probability$(s, q, \widetilde{\eta}, \widetilde{\Lambda}, \epsilon')$ reports $\widetilde{\mathsf{acc}}_{\epsilon'}(\mu)$ such that*

$$|\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda) - \mathsf{acc}_{\epsilon'}(\mu)| \leq \epsilon.$$

Proof: By using Lemma 4.4, we have $d_{\mathrm{TV}}(\mathcal{D}_{\mathsf{sim}}(\tilde{\eta}, \tilde{\Lambda}), \mathcal{D}_{\mathsf{test}}) \leq \epsilon$. We next analyze the calculated quantity $\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda)$. Let $\widetilde{\mathcal{D}}$ denote the distribution over $\{0, 1\}^{s \times q} \times A^s \times ([0, 1]^A)^q$ such that $\widetilde{\mathcal{D}}(M, a, t)$ is the probability that $\mathtt{Simulate}$ draws $(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_q) = t$, $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_s) = a$, and $\mathbf{M} = M$. Therefore, $\widetilde{\mathcal{D}}|_1 = \mathcal{D}_{\mathsf{sim}}(\tilde{\eta}, \tilde{\Lambda})$ and hence $\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda) = \mathbf{E}_{\mathbf{M} \sim \mathcal{D}_{\mathsf{sim}}(\tilde{\eta}, \tilde{\Lambda})}[\alpha_{\epsilon'}(\mathbf{M})]$.

This all means that

$$|\widetilde{\mathsf{acc}}_{\epsilon'}(\eta, \Lambda) - \mathsf{acc}_{\epsilon'}(\mu)| = \left| \mathop{\mathbf{E}}_{\mathbf{M} \sim \mathcal{D}_{\mathsf{sim}}(\tilde{\eta}, \tilde{\Lambda})} [\alpha_{\epsilon'}(\mathbf{M})] - \mathop{\mathbf{E}}_{\mathbf{M} \sim \mathcal{D}_{\mathsf{test}}} [\alpha_{\epsilon'}(\mathbf{M})] \right| \leq d_{\mathrm{TV}}(\mathcal{D}_{\mathsf{sim}}(\tilde{\eta}, \tilde{\Lambda}), \mathcal{D}_{\mathsf{test}}) \leq \epsilon$$

concluding the proof. ∎

5 Finding a weakly robust detailing and estimating its parameters

This section is devoted to finding a variable set defining a (weakly) robust partition, and then to estimating its weight and type distributions. The final estimation algorithm will make all its

queries deploying the algorithms developed here.

5.1 Estimating the index of a detailing

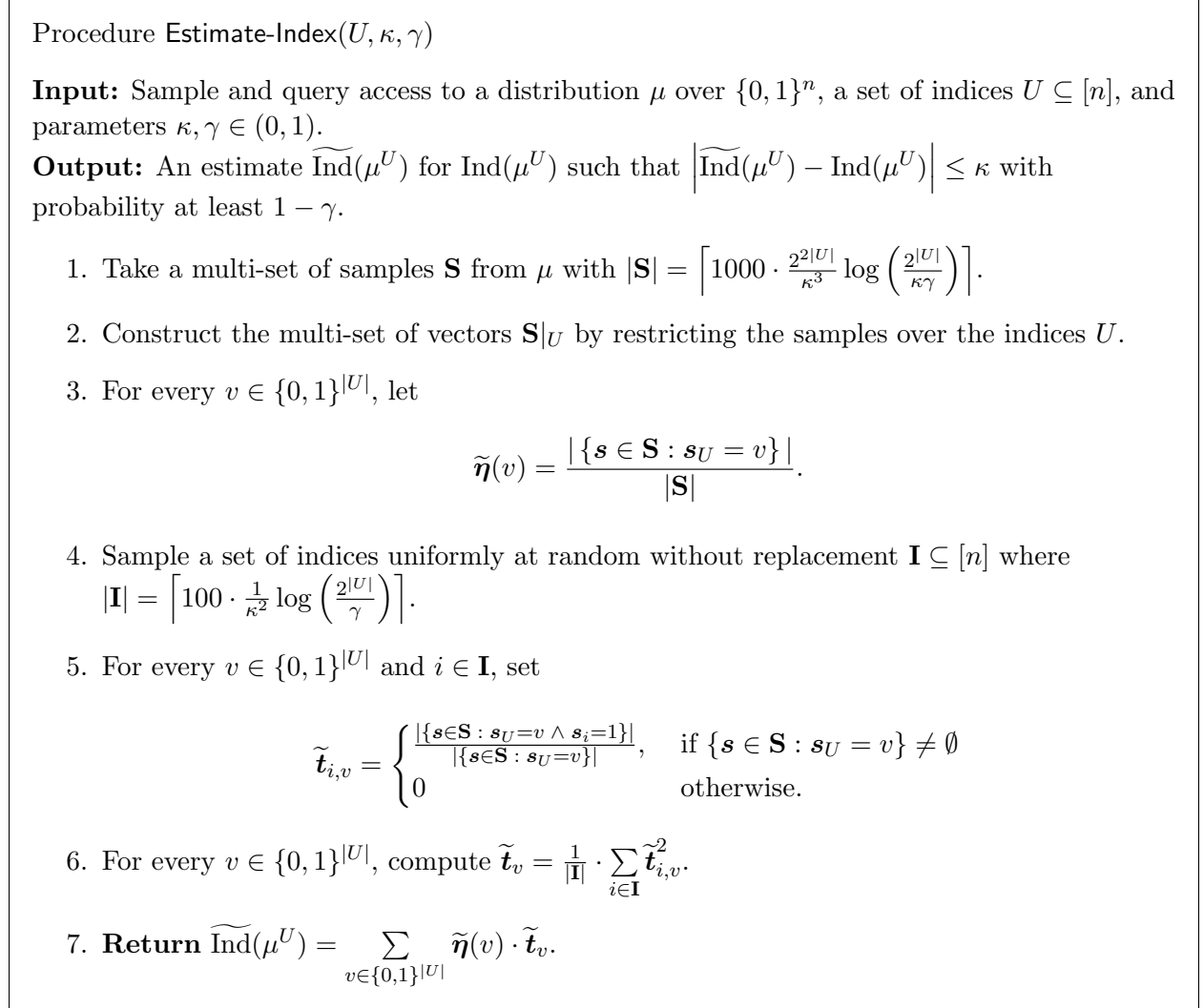In this section, we describe and analyze an algorithm for estimating the index of a detailing (Figure 3).

Procedure Estimate-Index(U, κ, γ)

Input: Sample and query access to a distribution μ over $\{0,1\}^n$, a set of indices $U \subseteq [n]$, and parameters $\kappa, \gamma \in (0,1)$.
Output: An estimate $\widetilde{\mathrm{Ind}}(\mu^U)$ for $\mathrm{Ind}(\mu^U)$ such that $\left| \widetilde{\mathrm{Ind}}(\mu^U) - \mathrm{Ind}(\mu^U) \right| \leq \kappa$ with probability at least $1 - \gamma$.

1. Take a multi-set of samples $\mathbf{S}$ from μ with $|\mathbf{S}| = \left\lceil 1000 \cdot \frac{2^{2|U|}}{\kappa^3} \log\left(\frac{2^{|U|}}{\kappa\gamma} \right) \right\rceil$.

2. Construct the multi-set of vectors $\mathbf{S}|_U$ by restricting the samples over the indices U.

3. For every $v \in \{0,1\}^{|U|}$, let

$$\widetilde{\boldsymbol{\eta}}(v) = \frac{\left| \{ \boldsymbol{s} \in \mathbf{S} : \boldsymbol{s}_U = v \} \right|}{|\mathbf{S}|}.$$

4. Sample a set of indices uniformly at random without replacement $\mathbf{I} \subseteq [n]$ where $|\mathbf{I}| = \left\lceil 100 \cdot \frac{1}{\kappa^2} \log\left(\frac{2^{|U|}}{\gamma} \right) \right\rceil$.

5. For every $v \in \{0,1\}^{|U|}$ and $i \in \mathbf{I}$, set

$$\widetilde{\boldsymbol{t}}_{i,v} = \begin{cases} \frac{|\{\boldsymbol{s} \in \mathbf{S} \, : \, \boldsymbol{s}_U = v \, \wedge \, \boldsymbol{s}_i = 1\}|}{|\{\boldsymbol{s} \in \mathbf{S} \, : \, \boldsymbol{s}_U = v\}|}, & \text{if } \{ \boldsymbol{s} \in \mathbf{S} : \boldsymbol{s}_U = v \} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

6. For every $v \in \{0,1\}^{|U|}$, compute $\widetilde{\boldsymbol{t}}_v = \frac{1}{|\mathbf{I}|} \cdot \sum_{i \in \mathbf{I}} \widetilde{\boldsymbol{t}}_{i,v}^2$.

7. **Return** $\widetilde{\mathrm{Ind}}(\mu^U) = \sum_{v \in \{0,1\}^{|U|}} \widetilde{\boldsymbol{\eta}}(v) \cdot \widetilde{\boldsymbol{t}}_v$.

Figure 3: Description of the Estimate-Index procedure.

We will prove that with high probability, the estimate of the index of μ^U returned by the algorithm Estimate-Index is indeed close to $\mathrm{Ind}(\mu^U)$.

Lemma 5.1. *Consider Estimate-Index(U, κ, γ) as described in Figure 3. Given a subset of indices $U \subseteq [n]$ and parameters $\kappa, \gamma \in (0,1)$ as input, the procedure makes at most $O\left(\frac{2^{2|U|}}{\kappa^5} \log^2\left(\frac{2^{|U|}}{\kappa\gamma} \right) \right)$ queries, and outputs $\widetilde{\mathrm{Ind}}(\mu^U)$ such that $|\widetilde{\mathrm{Ind}}(\mu^U) - \mathrm{Ind}(\mu^U)| \leq \kappa$ holds with probability at least $1 - \gamma$.*

Fix $v \in \{0,1\}^{|U|}$ and let $\eta(v) \overset{\text{def}}{=} \mu^U|_2(v) = \mathbf{Pr}_{\boldsymbol{x} \sim \mu}[\boldsymbol{x}_U = v]$. In addition, for $i \in [n]$ let

$$t_{i,v} = \begin{cases} \mathbf{Pr}_{\boldsymbol{w} \sim \mu^U|_1^{2:v}}[\boldsymbol{w}_i = 1], & \text{if } \mu^U|_2(v) \neq 0 \\ 0, & \text{otherwise} \end{cases}.$$

For $v \in \{0,1\}^{|U|}$, let $t_v = \frac{1}{n} \sum_{i=1}^{n} t_{i,v}^2$. Observe that the index of the detailing μ^U can be expressed as

$$\text{Ind}(\mu^U) = \sum_{v \in \{0,1\}^{|U|}} \eta(v) \cdot t_v.$$

Let $r = 2^{|U|}$, and define

$$J \overset{\text{def}}{=} \{v \in \{0,1\}^{|U|} : \eta(v) \geq \kappa/5r\}.$$

Definition 5.2 (Definition of the event $\mathcal{E}^*$). The event $\mathcal{E}^*$ is defined as the intersection of the following two events.

$\mathcal{E}_1$ (Approximating $\eta(v)$'s): For every $v \in \{0,1\}^{|U|}$, if $v \in J$ then $|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \leq \frac{\kappa}{10r}$, and if $v \notin J$ then $\widetilde{\boldsymbol{\eta}}(v) \leq \frac{3\kappa}{10r}$.

$\mathcal{E}_2$ (Approximating t_v's): For every $v \in J$, $|\widetilde{\boldsymbol{t}}_v - t_v| \leq \frac{6\kappa}{10}$.

Lemma 5.3. *Event $\mathcal{E}^*$ holds with probability at least $1 - \gamma$.*

For the proof of the above lemma we need the following two lemmas (in fact the next lemma is also used *inside* the proof of the following one).

Lemma 5.4 (Approximating $\eta(v)$'s). *Consider $v \in \{0,1\}^{|U|}$. The following hold with probability at least $1 - \frac{\kappa\gamma}{30r}$:*

    (i) If $\eta(v) \geq \frac{\kappa}{5r}$ then $|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \leq \frac{\kappa}{10r}$.

    (ii) If $\eta(v) < \frac{\kappa}{5r}$ then $\widetilde{\boldsymbol{\eta}}(v) \leq \frac{3\kappa}{10r}$.

Proof: For $\boldsymbol{s} \in \mathbf{S}$, we let $\boldsymbol{\chi}_s$ be the indicator random variable for $\boldsymbol{s}_U = v$ and note that $\mathbf{E}[\boldsymbol{\chi}_s] = \eta(v)$. Note that $\widetilde{\boldsymbol{\eta}}(v) = \sum_{s \in \mathbf{S}} \boldsymbol{\chi}_s$. Applying an additive Chernoff bound, we have that

$$\mathbf{Pr}\left[|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \geq \frac{\kappa}{10r}\right] \leq 2\exp\left(\frac{-2\kappa^2}{100r^2} \cdot |\mathbf{S}|\right) = 2\exp\left(\frac{-2\kappa^2}{100r^2} \cdot \frac{1000r^2}{\kappa^3} \log\left(\frac{r}{\kappa\gamma}\right)\right) < \frac{\kappa\gamma}{30r}.$$

Note that this covers both cases $v \in J$ and $v \notin J$. ∎

Lemma 5.5. *Consider $v \in J$. Then $|\widetilde{\boldsymbol{t}}_v - t_v| \leq \frac{6\kappa}{10}$ holds with probability at least $1 - \frac{2\gamma}{3r}$.*

Let us define $\boldsymbol{t}'_v$ as $\frac{1}{n} \sum_{i \in [n]} \widetilde{\boldsymbol{t}}_{i,v}^2$, where $v \in \{0,1\}^U$. We prove Lemma 5.5 with the help of the two following lemmas.

Lemma 5.6. *For any $v \in \{0,1\}^U$, $|\widetilde{\boldsymbol{t}}_v - \boldsymbol{t}'_v| \leq \frac{\kappa}{10}$ holds with probability at least $1 - \gamma/3r$.*

Proof: Note that by taking expectation over the set of indices $\mathbf{I}$, we have $\mathbf{E_I}\left[\widetilde{t}_v\right] = \frac{1}{n}\sum_{i=1}^n \widetilde{t}_{i,v}^2 = t'_v$. Applying Hoeffding's inequality for sampling without replacement (Lemma 2.33), we have

$$
\begin{aligned}
\mathbf{Pr}\left[|\widetilde{t}_v - t'_v| > \frac{\kappa}{10}\right] &= \mathbf{Pr}\left[\left|\frac{1}{|\mathbf{I}|}\sum_{i\in\mathbf{I}}\widetilde{t}_{i,v}^2 - \frac{1}{n}\sum_{i=1}^n \widetilde{t}_{i,v}^2\right| > \frac{\kappa}{10}\right] \\
&\leq 2\exp\left(-2\cdot\frac{\kappa^2}{100}\cdot|\mathbf{I}|\right) \\
&\leq 2\exp\left(-2\cdot\frac{\kappa^2}{100}\cdot\frac{100}{\kappa^2}\log\left(\frac{r}{\gamma}\right)\right) \\
&< \gamma/3r.
\end{aligned}
$$

∎

Lemma 5.7. *For any $v \in J$, $|t'_v - t_v| \leq \frac{5\kappa}{10}$ holds with probability at least $1 - \gamma/3r$.*

To prove the above lemma, we need the following definition of the notion of a variable being bad or good with respect to $v \in \{0,1\}^U$, and a claim about the probability of a variable being bad.

Definition 5.8. A variable $i \in [n]$ is said to be *bad* for $v \in \{0,1\}^U$ if $|\widetilde{t}_{i,v}^2 - t_{i,v}^2| > \frac{\kappa}{5r}$. Otherwise, i is said to be *good* for v.

Lemma 5.9. *For any fixed $i \in [n]$ and $v \in J$, i is bad for v with probability at most $\kappa\gamma/15r$.*

Proof: We first argue that

$$\mathbf{Pr}\left[|\widetilde{t}_{i,v} - t_{i,v}| \leq \frac{\kappa}{10}\right] \geq 1 - \frac{\kappa\gamma}{15r}.$$

This will imply that

$$\mathbf{Pr}\left[|\widetilde{t}_{i,v}^2 - t_{i,v}^2| \leq \frac{\kappa}{5}\right] \geq 1 - \frac{\kappa\gamma}{15r},$$

due to the fact that $|a^2 - b^2| \leq 2\cdot|a-b|$ when $a,b \in (0,1)$.

Let $\mathbf{S} = \{s^j\}_{j\in[|\mathbf{S}|]}$ be the multi-set of samples drawn in Step (1) of the procedure. For every $j \in [|\mathbf{S}|]$, consider an indicator random variable $\boldsymbol{\chi}_{j,v,i}$ such that $\boldsymbol{\chi}_{j,v,i} = 1$ if and only if $s^j_U = v$ and $s^j_i = 1$. Let $\boldsymbol{\chi}_{v,i} = \sum_{j\in[|\mathbf{S}|]}\boldsymbol{\chi}_{j,v,i}$. From the definition of $\widetilde{t}_{i,v}$ from Figure 3 and the definition of $\boldsymbol{\chi}_{v,i}$, we have

$$\widetilde{t}_{i,v} = \frac{|\{s \in \mathbf{S} : s_U = v \ \wedge\ s_i = 1\}|}{|\{s \in \mathbf{S} : s_U = v\}|} = \frac{\boldsymbol{\chi}_{v,i}}{|\{s \in \mathbf{S} : s|_U = v\}|}.$$

From the definition of $\widetilde{\boldsymbol{\eta}}(v)$ in Figure 3, this means that

$$\widetilde{t}_{i,v} = \frac{\boldsymbol{\chi}_{v,i}}{\widetilde{\boldsymbol{\eta}}(v)\cdot|\mathbf{S}|}. \tag{7}$$

In addition, note that for any $z \in \mathbb{N}$:

$$
\begin{aligned}
\mathbf{E}\left[\widetilde{t}_{i,v} \mid \widetilde{\boldsymbol{\eta}}(v)\cdot|\mathbf{S}| = z\right] = \mathbf{E}\left[\frac{\boldsymbol{\chi}_{v,i}}{\widetilde{\boldsymbol{\eta}}(v)\cdot|\mathbf{S}|} \,\middle|\, \widetilde{\boldsymbol{\eta}}(v)\cdot|\mathbf{S}| = z\right] &= \frac{z\cdot\mathbf{Pr}_{\boldsymbol{w}\sim\mu^U|_1^{w_U=v}}[\boldsymbol{w}_i = 1]}{z} \\
&= \mathbf{Pr}_{\boldsymbol{w}\sim\mu^U|_1^{w_U=v}}[\boldsymbol{w}_i = 1] = t_{i,v}.
\end{aligned}
$$

34

For $s \in \mathbf{S}$, let $\mathbf{Y}_{s,v,i}$ be the random variable defined as $\mathbf{Y}_{s,v,i} = \frac{\boldsymbol{\chi}_{s,v,i}}{\widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}|}$. Let $\mathbf{Y}_{v,i} = \sum_{s \in \mathbf{S}} \mathbf{Y}_{s,v,i}$. Since $\boldsymbol{\chi}_{s,v,i} \in \{0,1\}$, $\mathbf{Y}_{s,v,i} \in \left\{0, \frac{1}{\widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}|}\right\}$. We would like to apply the Hoeffding bound (Lemma 2.32) to bound $\mathbf{Y}_{v,i}$. But $\widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}|$ is a random variable, so instead we apply the Hoeffding bound conditioned on $\widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}| = z$, where $z \in \mathbb{N}$. Note that, when $\widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}| = z$, the number of s such that $\mathbf{Y}_{s,v,i} \neq 0$ is at most z. Hence,

$$\mathbf{Pr}\left[|\widetilde{\boldsymbol{t}}_{i,v} - t_{i,v}| > \frac{\kappa}{10} \,\middle|\, \widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}| = z\right] \leq 2\exp\left(-\frac{2(\kappa/10)^2}{z \cdot (1/z)^2}\right) = 2\exp\left(-\frac{2\kappa^2 z}{100}\right).$$

As $v \in J$, it holds that $\eta(v) \geq \frac{\kappa}{5r}$. From Lemma 5.4, with probability at least $1 - \kappa\gamma/30r$, it holds that $|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \leq \kappa/10r$, and in particular $\widetilde{\boldsymbol{\eta}}(v) \cdot |\mathbf{S}| \geq \frac{\kappa}{10r} \cdot |\mathbf{S}|$.

Hence,

$$\mathbf{Pr}\left[|\widetilde{\boldsymbol{t}}_{i,v} - t_{i,v}| > \frac{\kappa}{10}\right]$$

$$\leq \mathbf{Pr}\left[\widetilde{\boldsymbol{\eta}}(v)|\mathbf{S}| < \frac{\kappa}{10r} \cdot |\mathbf{S}|\right] + \mathbf{Pr}\left[\widetilde{\boldsymbol{\eta}}(v)|\mathbf{S}| \geq \frac{\kappa}{10} \cdot |\mathbf{S}|\right] \cdot \mathbf{Pr}\left[|\widetilde{\boldsymbol{t}}_{i,v} - t_{i,v}| > \frac{\kappa}{10} \,\middle|\, \widetilde{\boldsymbol{\eta}}(v)|\mathbf{S}| \geq \frac{\kappa}{10r} \cdot |\mathbf{S}|\right]$$

$$\leq \frac{\kappa\gamma}{30r} + 1 \cdot 2\exp\left(-2 \cdot \frac{\kappa^2 \cdot (\kappa/10r) \cdot |\mathbf{S}|}{100}\right)$$

$$\leq \frac{\kappa\gamma}{30r} + 2\exp\left(-2 \cdot \frac{\kappa^3}{1000r} \cdot \frac{1000r^2}{\kappa^3} \log\left(\frac{r}{\kappa\gamma}\right)\right) \leq \frac{\kappa\gamma}{15r}.$$

$\blacksquare$

Proof of Lemma 5.7: For $i \in [n]$, let $\mathbf{X}_i$ denote the random variable defined as

$$\mathbf{X}_i = \begin{cases} 1 & i \text{ is bad for } v \\ 0 & otherwise \end{cases}$$

Let $\mathbf{X} = \frac{1}{n}\sum_{i \in [n]} \mathbf{X}_i$, which denotes the fraction of variables that are bad for v. By Lemma 5.9, $\mathbf{E}[\mathbf{X}_i] \leq \frac{\kappa\gamma}{15r}$. Hence, $\mathbf{E}[\mathbf{X}] \leq \frac{\kappa\gamma}{15r}$. Using Markov Inequality, $\mathbf{Pr}[\mathbf{X} \geq \kappa/5] \leq \frac{\gamma}{3r}$. So, with probability $1 - \gamma/3r$, there are at most $\kappa n/5$ bad variables for v.

Note that,

$$|\boldsymbol{t}'_v - t_v| = \left|\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{t}}_{i,v}^2 - \frac{1}{n}\sum_{i=1}^{n} t_{i,v}^2\right| \leq \frac{1}{n}\sum_{i=1}^{n} \left|\widetilde{\boldsymbol{t}}_{i,v}^2 - t_{i,v}^2\right|.$$

For every variable $i \in [n]$ that is not bad for v, $\left|\widetilde{\boldsymbol{t}}_{i,v}^2 - t_{i,v}^2\right| \leq \frac{\kappa}{5}$. For every bad variable $i \in [n]$, $\left|\widetilde{\boldsymbol{t}}_{i,v}^2 - t_{i,v}^2\right|$ can be trivially upper bounded by 1. As there are there are at most $\kappa n/5$ bad variables for v with probability $1 - \gamma/3r$, we have $|\boldsymbol{t}'_v - t_v| \leq 2\kappa/5 < 6\kappa/10$ with the same probability. $\blacksquare$

Proof of Lemma 5.5: Considering Lemma 5.6 and Lemma 5.7, a use of the triangle inequality and a union bound complete the proof. $\blacksquare$

Proof of Lemma 5.3: Event $\mathcal{E}_1$ happens with probability at least $1 - \frac{\gamma}{3}$ by a union bound of Lemma 5.4 over all $v \in \{0,1\}^{|U|}$. Event $\mathcal{E}_2$ happens with probability at least $1 - \frac{2\gamma}{3}$ by a union bound

35

of Lemma 5.5 over all $v \in J$. A final union bound implies that event $\mathcal{E}^*$ occurs with probability at least $1 - \gamma$, as required. ∎

Now we are ready to prove Lemma 5.1.

Proof of Lemma 5.1: Conditioned on $\mathcal{E}^*$ (which by Lemma 5.3 holds with probability at least $1 - \gamma$) we have that, for every $v \in J$, $|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \leq \frac{\kappa}{10r}$ and $|\widetilde{\boldsymbol{t}}_v - t_v| \leq \frac{6\kappa}{10}$ hold. Hence,

$$
\begin{aligned}
\sum_{v \in J} |\widetilde{\boldsymbol{\eta}}(v)\widetilde{\boldsymbol{t}}_v - \eta(v)t_v| &\leq \sum_{v \in J} \widetilde{\boldsymbol{\eta}}(v)\left|\widetilde{\boldsymbol{t}}_v - t_v\right| + \sum_{v \in J} t_v |\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \\
&\leq \sum_{v \in J} \widetilde{\boldsymbol{\eta}}(v) \cdot \frac{6\kappa}{10} + \sum_{v \in J} t_v \cdot \frac{\kappa}{10r}
\end{aligned}
$$

Note that $\sum_{v \in J} \widetilde{\boldsymbol{\eta}}(v) \leq 1$ and $t_v \leq 1$ for each $v \in J$. So,

$$
\sum_{v \in J} |\widetilde{\boldsymbol{\eta}}(v)\widetilde{\boldsymbol{t}}_v - \eta(v)t_v| \leq \frac{7\kappa}{10}. \tag{8}
$$

For $v \notin J$, we have that

$$
|\widetilde{\boldsymbol{\eta}}(v)\widetilde{\boldsymbol{t}}_v - \eta_v t_v| \leq \max\{\widetilde{\boldsymbol{\eta}}(v)\widetilde{\boldsymbol{t}}_v, \eta(v)t_v\} \leq \max\{\widetilde{\boldsymbol{\eta}}(v), \eta(v)\} \leq \frac{3\kappa}{10r}.
$$

Therefore,

$$
\sum_{v \in \{0,1\}^{|U|} \setminus J} |\widetilde{\boldsymbol{\eta}}(v)\widetilde{\boldsymbol{t}}_v - \eta(v)t_v| \leq \frac{3\kappa}{10}. \tag{9}
$$

From Equation (8) and Equation (9) we have:

$$
\begin{aligned}
\left| \sum_{v \in \{0,1\}^{|U|}} \widetilde{\boldsymbol{\eta}}(v)\widetilde{\boldsymbol{t}}_v - \mathrm{Ind}(\xi^U) \right| &\leq \left| \sum_{v \in J} \widetilde{\boldsymbol{\eta}}(v) \cdot \widetilde{\boldsymbol{t}}_v - \eta(v)t_v \right| + \left| \sum_{v \in \{0,1\}^{|U|} \setminus J} \widetilde{\boldsymbol{\eta}}(v) \cdot \widetilde{\boldsymbol{t}}_v - \eta(v)t_v \right| \\
&\leq \frac{7\kappa}{10} + \frac{3\kappa}{10} = \kappa.
\end{aligned}
$$

For the query complexity, note that the total number of queries that the procedure makes to the samples is at most $|\mathbf{S}| \cdot (|\mathbf{I}| + |U|) = O\left(\frac{2^{2|U|}}{\kappa^5} \log^2\left(\frac{2^{|U|}}{\kappa\gamma}\right)\right)$. This completes the proof. ∎

5.2 Verifying and searching for a weakly robust detailing

In this section we will show how to obtain a weakly robust detailing with respect to a set of variables. We first describe a procedure that will be used as a single step towards this goal.

Lemma 5.10. *Fix $\gamma, \delta \in (0, 1)$, $U \subseteq [n]$, $k \in [n]$. Consider Test-Weakly-Robust-Detailing(δ, k, γ, U) as described in Figure 4. With probability at least $1 - \gamma$, the algorithm either accepts U or reports some $U' \subseteq [n] \setminus U$ such that $|U'| \leq k$, satisfying the following:*

1. *If μ^U is not (δ, k)-weakly robust, then Test-Weakly-Robust-Detailing does not accept and outputs some U';*

36

Procedure Test-Weakly-Robust-Detailing(δ, k, γ, U)

Input: Sample and query access to a distribution μ over $\{0,1\}^n$, $U \subseteq [n]$, parameters $k \in [n]$ and $\gamma \in (0,1)$.

Output: Either **Accept** or a set $U' \subseteq [n] \setminus U$ with $|U'| \leq k$.

1. Set $\widetilde{\mathrm{Ind}}(\mu^U) = \mathsf{Estimate\text{-}Index}(U, \delta/10, \gamma/3)$.

2. Let $\mathcal{L} = \{U' \subseteq [n] \setminus U : |U'| \leq k\}$.

3. Sample $r = \left\lceil \frac{\log(3/\gamma)}{\delta} \right\rceil$ sets $\mathbf{U}'_1, \ldots, \mathbf{U}'_r \in \mathcal{L}$ uniformly at random.

4. For $i = 1$ to r do:

   (a) Set $\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}'_i}) = \mathsf{Estimate\text{-}Index}(U \cup \mathbf{U}'_i, \delta/10, \gamma/3r)$.

   (b) If $\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}'_i}) - \widetilde{\mathrm{Ind}}(\mu^U) > 7\delta/10$, **Return** $\mathbf{U}'_i$.

5. **Return Accept**.

Figure 4: A description of the Test-Weakly-Robust-Detailing procedure.

2. *If Test-Weakly-Robust-Detailing outputs some U', then $\mathrm{Ind}(\mu^{U \cup U'}) > \mathrm{Ind}(\mu^U) + 4\delta/10$.*

The total number of queries that the algorithm makes is at most $O\left(\frac{2^{2(|U|+k)}}{\delta^6} \cdot \log^3\left(\frac{2^{|U|+k}}{\delta\gamma} \right) \right)$.

Proof: We start by proving (1). By Lemma 5.1, we have that with probability at least $1 - \gamma/3$, $|\widetilde{\mathrm{Ind}}(\mu^U) - \mathrm{Ind}(\mu^U)| \leq \delta/10$. In addition, since μ^U is not (δ, k)-weakly robust, we have that for at least $\delta|\mathcal{L}|$ of the sets $U' \in \mathcal{L}$, it holds that $\mathrm{Ind}(\mu^{U \cup U'}) - \mathrm{Ind}(\mu^U) > \delta$. Therefore, with probability at least $1 - \gamma/3$, one such set $\mathbf{U}^*$ is sampled in Step 3 of the algorithm. By applying Lemma 5.1 and a union bound, we have that with probability at least $1 - \gamma/3$, for all $i \in [r]$ it holds that $|\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}'_i}) - \mathrm{Ind}(\mu^{U \cup \mathbf{U}'_i})| \leq \delta/10$. Therefore, in particular,

$$\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}^*}) - \widetilde{\mathrm{Ind}}(\mu^U) \geq \mathrm{Ind}(\mu^{U \cup \mathbf{U}^*}) - \mathrm{Ind}(\mu^U) - \frac{2\delta}{10} \geq \frac{8\delta}{10},$$

and $\mathbf{U}^*$ (or some other set) will be returned by the algorithm. A union bound over the above three events completes the proof of this item.

To prove (2), note that when some $\mathbf{U}'$ is returned, we have that $\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}'}) - \widetilde{\mathrm{Ind}}(\mu^U) > 7\delta/10$. By Lemma 5.1, similarly to item (1), with probability at least $1 - 2\gamma/3$ we have $|\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}'}) - \mathrm{Ind}(\mu^{U \cup \mathbf{U}'})| \leq \delta/10$ and $|\widetilde{\mathrm{Ind}}(\mu^U) - \mathrm{Ind}(\mu^U)| \leq \delta/10$. Conditioned on this, if $\mathrm{Ind}(\mu^{U \cup \mathbf{U}'}) - \mathrm{Ind}(\mu^U) \leq 4\delta/10$, then we have

$$\widetilde{\mathrm{Ind}}(\mu^{U \cup \mathbf{U}'}) - \widetilde{\mathrm{Ind}}(\mu^U) \leq \frac{4\delta}{10} + \frac{2\delta}{10} < \frac{7\delta}{10},$$

which is a contradiction to $\mathbf{U}'$ being reported by the algorithm.

To finish the proof we bound the total number of queries that the algorithm makes. Note that the algorithm uses one call to Estimate-Index with U and the rest of the r calls to Estimate-Index

37

with a set $U \cup U'$ which can be of size at most $|U| + k$. Therefore, the total number of queries the algorithm makes is at most $O\left(\frac{2^{2(|U|+k)}}{\delta^6} \cdot \log^3\left(\frac{2^{|U|+k}}{\delta\gamma}\right)\right)$. ■

Now we leverage Test-Weakly-Robust-Detailing to design a procedure, Find-Weakly-Robust-Detailing (see Figure 5), which finds a weakly robust detailing μ^U with respect to some $U \subseteq [n]$. This procedure along with the parameter estimation procedure in the next section serve as the main information-gathering mechanism in our estimation algorithm.

Procedure Find-Weakly-Robust-Detailing(δ, k, γ)

Input: Sample and query access to a distribution μ over $\{0,1\}^n$, $k \in [n]$ and parameters $\delta, \gamma \in (0,1)$.
Output: $U \subseteq [n]$ such that with probability at least $1 - \gamma$, μ^U is (δ, k)-weakly robust.

1. Set $U = \emptyset$ and $\ell = 1$.

2. While $\ell \leq \lceil \frac{10}{4\delta} \rceil$ do:

   (a) Set $\mathbf{Z} = $ Test-Weakly-Robust-Detailing$(\delta, k, 2\gamma\delta/10, U)$

   (b) If $\mathbf{Z} = \mathbf{Accept}$, then **Return** U.

   (c) Otherwise,

   - $U' \leftarrow \mathbf{Z}$.
   - Set $U = U \cup U'$ and $\ell = \ell + 1$.

3. **Return Fail**

Figure 5: A description of the Find-Weakly-Robust-Detailing procedure.

Lemma 5.11. *Consider Find-Weakly-Robust-Detailing(δ, k, γ) as described in Figure 5. With probability at least $1 - \gamma$, the procedure outputs $U \subset [n]$ such that $|U| = O(k/\delta)$ and μ^U is (δ, k)-weakly robust. The total number of queries that the procedure makes is at most $O\left(\frac{2^{5k(1/\delta+1)}}{\delta^7} \log^3\left(\frac{2^{k(1/\delta+1)}}{\gamma\delta}\right)\right)$.*

Proof: Note that since $\ell \leq \lceil 10/4\delta \rceil$, by a union bound we have that with probability at least $1 - \gamma$ all the calls for Test-Weakly-Robust-Detailing$(\delta, k, 2\gamma\delta/10, U)$ satisfy the conclusions of Lemma 5.10. Conditioned on this event, whenever in Step (2) the detailing μ^U was not (δ, k)-weakly robust we obtained a set U' so that $\text{Ind}(\mu^{U'}) - \text{Ind}(\mu^U) > 4\delta/10$ (and in particular did not return U).

In addition, conditioned on the above event, since the index of a detailing is always between 0 and 1 it is not possible to pass through all $\lceil \frac{10}{4\delta} \rceil$ iterations of Step (2) and reach the failure mode of Step (3). Hence, eventually the algorithm will return a set U in Step (2b), such that μ^U is a (δ, k)-weakly robust detailing.

To bound the number of queries note that since the algorithm starts with $U = \emptyset$, and by the guarantees of Test-Weakly-Robust-Detailing, at any of the ℓ iterations of the algorithm the size of the input U to Test-Weakly-Robust-Detailing is at most $k(\frac{10}{4\delta} + 1) = O(k/\delta)$. Therefore the total number of queries is at most $O\left(\frac{2^{5k(1/\delta+1)}}{\delta^7} \log^3\left(\frac{2^{k(1/\delta+1)}}{\gamma\delta}\right)\right)$. ■

5.3 Estimation of detailing parameters

In this section we will design a procedure, Estimate-Parameters (see Figure 6), for estimating the parameters of a detailing μ^U with respect to some $U \subseteq [n]$ admitting type distribution Λ and weight distribution $\eta = \mu^U|_2$. Our main goal for this section is to prove the following.

Lemma 5.12. *Let μ be a distribution over $\{0,1\}^n$, $U \subseteq [n]$ and $\kappa, \gamma \in (0,1)$ be parameters. Then the procedure Estimate-Parameters(μ, U, κ, γ) (see Figure 6) outputs a pair of distributions $(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\Lambda}})$ such that with probability at least $1 - \gamma$, $d_{\mathrm{TV}}(\widetilde{\boldsymbol{\eta}}, \eta) \leq \kappa$ and $d_{\mathrm{EM}}^\eta(\widetilde{\boldsymbol{\Lambda}}, \Lambda) \leq \kappa$, where $\eta = \mu^U$ is the weight distribution of μ^U and Λ is the type distribution of μ^U. The algorithm makes at most $O\left(\frac{2^{2|U|}}{\kappa^{2^{|U|+1}+5}} \cdot \log^2\left(\frac{2^{|U|}}{\gamma\kappa^{2^{|U|}}}\right)\right)$ queries to the input.*

We start by setting the stage with some definitions and preliminary results. Consider some $\rho \in (0,1)$ that satisfies $\rho = 1/z$ for $z \in \mathbb{N}$. We use $\mathcal{R}$ to denote the set $\{0, \rho/10, \ldots, 1\}^{2^{|U|}}$ and note that $|\mathcal{R}| = (10/\rho + 1)^{2^{|U|}}$.

Definition 5.13. Fix $U \subseteq [n]$. We say that a vector $w \in [0,1]^{2^{|U|}}$ is a *ρ-approximation* of the type $t \in [0,1]^{2^{|U|}}$ of μ^U if $|w(v) - t(v)| \leq \rho$ for any $v \in \{0,1\}^{2^{|U|}}$ for which $\mu^U|_2(v) \geq \rho/2^{|U|}$.

Observation 5.14. *Let $\eta = \mu^U|_2$ and fix $i \in [n]$. Let $t_i \in [0,1]^{2^U}$ be the type of a variable i with respect to μ^U, and let $w \in [0,1]^{2^U}$ be a ρ-approximation of t_i. Then, $d_{\ell_1}^\eta(t_i, w) \leq 2\rho$.*

Proof: By definition of the η-weighted ℓ_1 distance,

$$
\begin{aligned}
d_{\ell_1}^\eta(t_i, w) &= \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\eta}[|t_i(\boldsymbol{v}) - w(\boldsymbol{v})|] \\
&= \sum_{v\in\{0,1\}^U : \eta(v)\geq\rho/2^{|U|}} \eta(v) \cdot |t_i(v) - w(v)| + \sum_{v\in\{0,1\}^U : \eta(v)<\rho/2^{|U|}} \eta(v) \cdot |t_i(v) - w(v)| \\
&\leq \rho + 2^{|U|} \cdot \frac{\rho}{2^{|U|}} = 2\rho.
\end{aligned}
$$

∎

Definition 5.15. Let $U \subseteq [n]$, S be a multi-set over $\{0,1\}^n$, and $\rho \in (0,1)$ be a parameter so that $1/\rho \in \mathbb{N}$. For each $v \in \{0,1\}^{|U|}$ and $i \in [n]$, let $\widehat{\alpha}_v^{(i)}$ to be the fraction of x's in S such that $x_U = v$ and $x_i = 1$, and let $\widetilde{\alpha}_v^{(i)}$ be the element in $\{0, \rho, 2\rho, \ldots, 1\}$ closest to $\widehat{\alpha}_v^{(i)}$. Let $\widehat{\alpha}^{(i)} = \langle\widehat{\alpha}_v^{(i)}\rangle_{v\in\{0,1\}^{|U|}}$ and $\widetilde{\alpha}^{(i)} = \langle\widetilde{\alpha}_v^{(i)}\rangle_{v\in\{0,1\}^{|U|}}$. We call $\widehat{\alpha}^{(i)}$ the *type of i according to S* and $\widetilde{\alpha}^{(i)}$ its *ρ-rounding*.

We also define the distribution $\widehat{\Lambda}$ as the distribution over $\mathcal{R} = \{0, \rho, 2\rho, \ldots, 1\}^{2^{|U|}}$ such that the probability of $a \in \mathcal{R}$ is the fraction of $i \in [n]$ such that $\widetilde{\alpha}^{(i)} = a$.

Definition 5.16 (ρ-Good Sample Set). A multi-set S over $\{0,1\}^n$ is said to be a *ρ-good sample set* with respect to μ^U, if for at least $(1 - \rho)n$ of the variables $i \in [n]$, $\widetilde{\alpha}^{(i)}$ is a ρ-approximation of the type t_i of the detailing μ^U.

Lemma 5.17. *Fix $U \subseteq [n]$, $\rho \in (0,1)$ and let μ^U be the detailing of μ with respect to U admitting type distribution Λ. If $S \subseteq \{0,1\}^n$ is a ρ-good sample set with respect to μ^U, then letting $\eta = \mu^U|_2$ we have $d_{\mathrm{EM}}^\eta(\Lambda, \widehat{\Lambda}) \leq 3\rho$.*

Procedure Estimate-Parameters(μ, U, κ, γ)

Input: Sample and query access to a distribution μ over $\{0,1\}^n$, $U \subseteq [n]$ and parameters $\kappa, \gamma \in (0,1)$.

Output: A pair of distributions $(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\Lambda}})$, where $\widetilde{\boldsymbol{\eta}}$ is a distribution on $A = \{0,1\}^{|U|}$ and $\widetilde{\boldsymbol{\Lambda}}$ is a distribution on $[0,1]^A$.

1. Set $\rho = \frac{1}{4 \cdot \lceil 1/\kappa \rceil}$ and $r = 2^{|U|}$. Also, let $\mathcal{R} = \{0, \rho/10, \ldots, 1\}^{2^{|U|}}$.

2. Take a multi-set of samples $\mathbf{S}$ from μ with $|\mathbf{S}| = \left\lceil 50 \cdot \frac{r^2}{\rho^3} \log \frac{r}{\rho\gamma} \right\rceil = O\left(\frac{2^{2|U|}}{\kappa^3} \log \frac{2^{|U|}}{\kappa\gamma} \right)$.

3. Construct the multi-set of vectors $\mathbf{S}|_U$ by restricting the samples over the indices U.

4. For every $v \in \{0,1\}^{|U|}$ let
$$\widetilde{\boldsymbol{\eta}}(v) = \frac{\left| \{ \boldsymbol{s} \in \mathbf{S} : \boldsymbol{s}|_U = v \} \right|}{|\mathbf{S}|}.$$

5. Sample a set of indices uniformly at random without replacement $\mathbf{I} \subseteq [n]$ where
$|\mathbf{I}| = \left\lceil 20 \cdot \frac{|\mathcal{R}|^2}{\rho^2} \log \frac{|\mathcal{R}|}{\gamma} \right\rceil$, noting that $|\mathcal{R}| = \left(\frac{10}{\rho} + 1 \right)^{2^{|U|}}$.

6. For every $v \in \{0,1\}^{|U|}$ and $i \in \mathbf{I}$, set

   (i) $\widehat{\boldsymbol{\alpha}}_{v,i} = \frac{|\{ \boldsymbol{s} \in \mathbf{S} : \, \boldsymbol{s}|_U = v \, \wedge \, \boldsymbol{s}_i = 1 \}|}{|\{ \boldsymbol{s} \in \mathbf{S} : \, \boldsymbol{s}|_U = v \}|}$ if $\{ \boldsymbol{s} \in \mathbf{S} : \boldsymbol{s}|_U = v \} \neq \emptyset$, and $\widehat{\boldsymbol{\alpha}}_{v,i} = 0$ otherwise.

   (ii) Round off $\widehat{\boldsymbol{\alpha}}_{v,i}$ to the nearest value in $\{0, \rho/10, \ldots, 1\}$, and denote this value by $\widetilde{\boldsymbol{\alpha}}_{v,i}$.

7. For each $i \in \mathbf{I}$, assign it the type vector $\widetilde{\boldsymbol{\alpha}}^{(i)} = \langle \widetilde{\boldsymbol{\alpha}}_{v,i} \rangle_{v \in \{0,1\}^{|U|}} \in \mathcal{R}$.

8. Construct a distribution $\widetilde{\boldsymbol{\Lambda}}$ over $\mathcal{R}$ such that for each $a \in \mathcal{R}$, we have:
$$\widetilde{\boldsymbol{\Lambda}}(a) = \frac{|\{ i \in \mathbf{I} : \widetilde{\boldsymbol{\alpha}}^{(i)} = a \}|}{|\mathbf{I}|}.$$

9. **Return** $(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\Lambda}})$.

Figure 6: A description of the Estimate-Parameters procedure.

Proof: As S is a ρ-good sample set, we know that for at least $(1-\rho)n$ of the variables $i \in [n]$, $\widetilde{\alpha}^{(i)}$ is a ρ-approximation of the type t_i. Let $W \subseteq [n]$ be the set of all such variables with $|W| \geq (1-\rho)n$. By Observation 5.14,
$$W \subseteq \left\{ i \in [n] : d_{\ell_1}^{\eta}(\widetilde{\alpha}^{(i)}, t_i) \leq 2\rho \right\}.$$

Consider the function $f : \left([0,1]^{2^{|U|}} \right)^2 \to [0,1]$ defined as follows.

$$f(a,b) = \begin{cases} \frac{|\{ i \in W : t_i = a \}|}{n} & \widetilde{\alpha}^{(i)} = b \\ 0 & otherwise \end{cases}.$$

40

From the definition of f and Observation 5.14, note that $f(a, b) \neq 0$ implies that $d^\eta_{\ell_1}(a, b) \leq 2\rho$. Consider a transfer function $T : \left([0, 1]^{2^{|U|}}\right)^2 \to [0, 1]$ from $\widehat{\Lambda}$ to Λ satisfying $T(a, b) \geq f(a, b)$ for all $a, b \in [0, 1]^{2^{|U|}}$. Note that such a T exists, since clearly $\sum_{a, b \in [0,1]^{2^{|U|}}} f(a, b) \leq 1$. Also note that $\sum_{a, b \in [0,1]^{2^{|U|}}} f(a, b) = 1 - |W|/n \geq 1 - \rho$, and hence $\sum_{a, b \in [0,1]^{2^{|U|}} : f(a,b)=0} T(a, b) \leq \rho$. Thus, we have

$$
\begin{aligned}
d^\eta_{\mathrm{EM}}(\widehat{\Lambda}, \Lambda) &\leq \mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim T} [d^\eta_{\ell_1}(\boldsymbol{a}, \boldsymbol{b})] \\
&= \sum_{a, b \in [0,1]^{2^{|U|}} : f(a,b) \neq 0} T(a, b) \cdot d^\eta_{\ell_1}(a, b) + \sum_{a, b \in [0,1]^{2^{|U|}} : f(a,b) = 0} T(a, b) \cdot d^\eta_{\ell_1}(a, b) \\
&\leq 2\rho \cdot \sum_{a, b \in [0,1]^{2^{|U|}} : f(a,b) \neq 0} T(a, b) + \sum_{a, b \in [0,1]^{2^{|U|}} : f(a,b) = 0} T(a, b) \cdot 1 \leq 3\rho,
\end{aligned}
$$

and the lemma follows. $\blacksquare$

Definition 5.18. Let $U \subseteq [n]$, S be a multi-set over $\{0, 1\}^n$, and $\rho \in (0, 1)$ where $1/\rho \in \mathbb{N}$ be a parameter. Let $\widehat{\alpha}^{(i)}$ be the type of i according to S and $\widetilde{\alpha}^{(i)}$ be its ρ-rounding over S, as per Definition 5.15. For each $a \in \mathcal{R}$, let f_a be the fraction of $i \in [n]$ such that $\widetilde{\alpha}^{(i)} = a$. Let $I \subseteq [n]$ be a multi-set of variables. For each $a \in \mathcal{R}$, let $\widetilde{f}_a$ be the fraction of $i \in I$ such that $\widetilde{\alpha}^{(i)} = a$.

The approximate type distribution $\widetilde{\Lambda}$ with respect to I and S is a distribution over $\mathcal{R}$ such that the probability of $a \in \mathcal{R}$ is the fraction of $i \in I$ such that $\widetilde{\alpha}^{(i)} = a$.

Definition 5.19 (ρ-Good Variable Set). Let S be a multi-set over $\{0, 1\}^n$. Using the notation of Definition 5.18, a multi-set I over $[n]$ is said to be ρ-*Good-Variable-Set* with respect to S if $|\widetilde{f}_a - f_a| \leq \rho/|\mathcal{R}|$ for each $a \in \mathcal{R}$.

Lemma 5.20. *Let S be a multi-set over $\{0, 1\}^n$ and $I \subseteq [n]$ be a ρ-good variable set with respect to S. Then, $d_{\mathrm{TV}}(\widehat{\Lambda}, \widetilde{\Lambda}) \leq \rho$.*

Proof: By definition of the distributions $\widehat{\Lambda}$ and $\widetilde{\Lambda}$,

$$
d_{\mathrm{TV}}(\widehat{\Lambda}, \widetilde{\Lambda}) \leq \sum_{a \in \mathcal{R}} |\widehat{\Lambda}(a) - \widetilde{\Lambda}(a)| = \sum_{a \in \mathcal{R}} |\widetilde{f}_a - f_a| \leq |\mathcal{R}| \cdot \frac{\rho}{|\mathcal{R}|} \leq \rho.
$$

$\blacksquare$

Lemma 5.21. *Let $\eta = \mu^U|_2$ for some $U \subseteq [n]$. Let $S \subseteq \{0, 1\}^n$ be a ρ-good-sample set and $I \subset [n]$ be a ρ-good variable set with respect to S. Then, $d^\eta_{\mathrm{EM}}(\Lambda, \widetilde{\Lambda}) \leq 4\rho$.*

Proof: From Lemma 5.17, we know that $d^\eta_{\mathrm{EM}}(\Lambda, \widehat{\Lambda}) \leq 3\rho$. Also, from Lemma 5.20, we have that $d_{\mathrm{TV}}(\widehat{\Lambda}, \widetilde{\Lambda}) \leq \rho$, and hence $d^\eta_{\mathrm{EM}}(\widehat{\Lambda}, \widetilde{\Lambda}) \leq \rho$. Using the triangle inequality we obtain $d^\eta_{\mathrm{EM}}(\Lambda, \widetilde{\Lambda}) \leq 4\rho$, completing the proof of the lemma. $\blacksquare$

5.3.1 Proof of Lemma of 5.12

Definition 5.22 (Definition of the event $\mathcal{E}^*$). The event $\mathcal{E}^*$ is defined as the intersection of the following three events:

$\mathcal{E}_{\widetilde{\boldsymbol{\eta}}}$: The output $\widetilde{\boldsymbol{\eta}}$ produced by Algorithm 6 satisfies $d_{\mathrm{TV}}(\widetilde{\boldsymbol{\eta}}, \eta) \leq \kappa$.

$\mathcal{E}_{\mathbf{S}}$: The multi-set of samples $\mathbf{S}$ taken in Step 2 of Algorithm 6 is a $\kappa/4$-good sample set.

$\mathcal{E}_{\mathbf{I}}$: The set of indices $\mathbf{I}$ taken in Step 5 of Algorithm 6 is a $\kappa/4$-good variable set.

We will prove Lemma 5.12 by a series of lemmas stated below. The lemmas themselves will be proved in Section 5.3.2.

Lemma 5.23. *Event $\mathcal{E}_{\widetilde{\boldsymbol{\eta}}}$ holds with probability at least $1 - \gamma/3$.*

Lemma 5.24. *Event $\mathcal{E}_{\mathbf{S}}$ holds with probability at least $1 - \gamma/3$.*

Lemma 5.25. *Event $\mathcal{E}_{\mathbf{I}}$ holds with probability at least $1 - \gamma/3$.*

Proof of Lemma 5.12: Assuming that Lemma 5.23, Lemma 5.24 and Lemma 5.25 hold, the event $\mathcal{E}^*$ holds with probability at least $1 - \gamma$ by a union bound.

When $\mathcal{E}^*$ holds, we know that $d_{\mathrm{TV}}(\widetilde{\boldsymbol{\eta}}, \eta) \leq \kappa$, the multi-set of samples $\mathbf{S}$ taken in Step 2 is a good sample set as well as that the set of indices $\mathbf{I}$ is a good variable set. Following Lemma 5.21, we have that $d_{\mathrm{EM}}^{\eta}(\Lambda, \widetilde{\boldsymbol{\Lambda}}) \leq 4\rho \leq \kappa$.

Note that since $|\mathcal{R}| = (10/\rho + 1)^{2^{|U|}}$ and $\rho = O(\kappa)$, the total number of queries the algorithm makes is at most $|\mathbf{S}| \cdot (|\mathbf{I}| + |U|) = O\left(\frac{2^{2|U|}}{\kappa^{2^{|U|+1}+5}} \cdot \log^2 \left(\frac{2^{|U|}}{\gamma \kappa^{2^{|U|}}} \right) \right)$. ∎

5.3.2 Proofs of Lemmas 5.23, 5.24 and 5.25

We state the following lemmas whose proofs are identical to the proofs of Lemma 5.4 and Lemma 5.5 respectively and are thus omitted.

Lemma 5.26. *Fix $U \subseteq [n]$ and $\rho, \gamma \in (0,1)$ and let $r = 2^{|U|}$. Consider the detailing μ^U and let $\eta = \mu^U|_2$ and $J = \{v \in \{0,1\}^{|U|} : \eta(v) \geq \rho/r\}$. For $v \in \{0,1\}^{|U|}$, we define the following event $\mathcal{E}_{\widetilde{\boldsymbol{\eta}}(v)}$: The value $\widetilde{\boldsymbol{\eta}}(v)$ defined in Step 4 of Estimate-Parameters satisfies the following.*

(i) If $v \in J$, $|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \leq \frac{\rho}{2r}$.

(ii) If $v \notin J$, $\widetilde{\boldsymbol{\eta}}(v) \leq \frac{3\rho}{2r}$.

Then the probability that $\mathcal{E}_{\widetilde{\boldsymbol{\eta}}(v)}$ holds is at least $1 - \frac{\rho\gamma}{6r}$.

Lemma 5.27. *Fix $U \subseteq [n]$ and $\rho, \gamma \in (0,1)$ and let $r = 2^{|U|}$. Consider the detailing μ^U and let $\eta = \mu^U|_2$ and $J = \{v \in \{0,1\}^{|U|} : \eta(v) \geq \rho/r\}$. For each $v \in J$ and $i \in [n]$, we let $\mathcal{E}_{\widehat{\boldsymbol{\alpha}}_{v,i}}$ be the event that the value $\widehat{\boldsymbol{\alpha}}_{v,i}$ satisfies $|\widehat{\boldsymbol{\alpha}}_{v,i} - t_i(v)| \leq \frac{\rho}{4}$. Then the probability of $\mathcal{E}_{\widehat{\boldsymbol{\alpha}}_{v,i}}$ is at least $1 - \frac{\rho\gamma}{3r}$.*

From Lemma 5.27 we obtain the following lemma.

Lemma 5.28. *Fix $i \in [n]$. Then $\widetilde{\boldsymbol{\alpha}}^{(i)}$ is a ρ-approximation of t_i with probability at least $1 - \rho\gamma/3$.*

Proof: Let $\mathcal{E}$ be the following event:

$$\mathcal{E}: \ |\widetilde{\boldsymbol{\alpha}}_v^{(i)} - t_i(v)| \leq \rho \text{ for each } v \in J.$$

If the event $\mathcal{E}$ holds, then $\widetilde{\boldsymbol{\alpha}}^{(i)}$ is a ρ-approximation of t_i. Hence, it remains to show that $\mathbf{Pr}[\mathcal{E}] \geq 1 - \gamma\rho/3$.

Consider any particular $v \in J$, i.e, $\eta(v) \geq \frac{\rho}{r}$. By Lemma 5.27, $|\widehat{\boldsymbol{\alpha}}_v^{(i)} - t_i(v)| \leq \frac{\rho}{4}$ with probability at least $1 - \rho\gamma/3r$. From the construction of $\widetilde{\boldsymbol{\alpha}}_v^{(i)}$, note that $|\widetilde{\boldsymbol{\alpha}}_v^{(i)} - \widehat{\boldsymbol{\alpha}}_v^{(i)}| \leq \rho/20$. By the triangle inequality, $|\widetilde{\boldsymbol{\alpha}}_v^{(i)} - t_i(v)| \leq 3\rho/10 < \rho$. By the union bound over all $v \in J$, event $\mathcal{E}$ holds with probability at least $1 - \rho\gamma/3$. ∎

Now we are ready to prove Lemma 5.23, Lemma 5.24 and Lemma 5.25.

Proof of Lemma 5.23: By Lemma 5.26 and a union bound we have that the following holds for all $v \in \{0,1\}^{|U|}$ with probability at least $1 - \rho\gamma/4 > 1 - \gamma/3$. If $v \in J$, $|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| \leq \frac{\rho}{2r}$, and otherwise $\widetilde{\boldsymbol{\eta}}(v) \leq \frac{3\rho}{2r}$. Conditioned on that,

$$d_{\mathrm{TV}}(\widetilde{\boldsymbol{\eta}}, \eta) = \frac{1}{2}\left(\sum_{v \in J}|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)| + \sum_{v \notin J}|\widetilde{\boldsymbol{\eta}}(v) - \eta(v)|\right) \leq \frac{1}{2}\left(\sum_{v \in J}\frac{\rho}{2r} + \sum_{v \notin J}\frac{3\rho}{2r}\right) \leq \rho.$$

As $\rho = \frac{1}{4\lceil 1/\kappa\rceil}$ the lemma follows. ∎

Proof of Lemma 5.24: Let $\mathbf{X}_i$ be a random variable such that

$$\mathbf{X}_i = \begin{cases} 1 & \widetilde{\boldsymbol{\alpha}}^{(i)} \text{ is not a } \rho\text{-approximation of } t_i \\ 0 & \text{otherwise} \end{cases}.$$

Let $\mathbf{X} = \frac{1}{n}\sum_{i \in [n]}\mathbf{X}_i$. From Lemma 5.28, $\mathbf{E}[\mathbf{X}_i] \leq \frac{\rho\gamma}{3}$, and hence $\mathbf{E}[\mathbf{X}] \leq \frac{\rho\gamma}{3}$. By Markov inequality, we have

$$\mathbf{Pr}\left[\mathbf{X} \geq \rho\right] \leq \frac{\mathbf{E}[\mathbf{X}]}{\rho} \leq \frac{\gamma}{3}.$$

Thus, with probability at least $1-\gamma/3$, for at least $(1-\rho)n$ variables $i \in [n]$, $\widetilde{\boldsymbol{\alpha}}^{(i)}$ is a ρ-approximation of t_i. Hence, $\mathbf{S}$ is a ρ-good sample set with probability at at least $1 - \gamma/3$. ∎

Proof of Lemma 5.25: Recall that $\widehat{\Lambda}$ be the type distribution over $\mathcal{R}$ with respect to $\mathbf{S}$ such that, for $a \in \mathcal{R}$,

$$\widehat{\Lambda}(a) = \frac{|\{i \in [n] : \widetilde{\boldsymbol{\alpha}}^{(i)} = a\}|}{n}.$$

Also, consider the distribution $\widetilde{\Lambda}$ reported by the algorithm. Note that $\widetilde{\Lambda}$ is the type distribution over $\mathcal{R}$ with respect to $\mathbf{I}$ and $\mathbf{S}$, i.e., for $a \in \mathcal{R}$,

$$\widetilde{\Lambda}(a) = \frac{|\{i \in \mathbf{I} : \widetilde{\boldsymbol{\alpha}}^{(i)} = a\}|}{|\mathbf{I}|}.$$

Consider a fixed $a \in \mathcal{R}$, and note that $\mathbf{E}_{\mathbf{I}}[\widetilde{\Lambda}(a)] = \widehat{\Lambda}(a)$. By using the Hoeffding bound (Lemma 2.33),

$$\mathbf{Pr}\left[|\widetilde{\Lambda}(a) - \widehat{\Lambda}(a)| \geq \frac{\rho}{|\mathcal{R}|}\right] \leq \frac{\gamma}{3|\mathcal{R}|}.$$

43

In the above, we have used $|\mathbf{I}| = \left\lceil 20 \cdot \frac{|\mathcal{R}|^2}{\rho^2} \log \frac{|\mathcal{R}|}{\gamma} \right\rceil$. Applying the union bound over all $a \in \mathcal{R}$, we can say that for all $a \in \mathcal{R}$, $|\widetilde{\mathbf{\Lambda}}(a) - \widehat{\Lambda}(a)| \leq \frac{\rho}{|\mathcal{R}|}$ with probability at least $1 - \gamma/3$. Since $\rho = \frac{1}{4\lceil 1/\kappa \rceil}$, this implies that the event $\mathcal{E}_{\mathbf{I}}$ holds with probability at least $1 - \gamma/3$. This concludes the proof of the lemma. ∎

6 The estimation algorithm

In this section we prove our main result, that index-invariant properties that admit tests whose number of queries is independent of n also admit such distance-estimation procedures.

Theorem 2. *Let $\mathcal{P}$ be an index-invariant property of distributions supported on $\{0,1\}^n$. If $\mathcal{P}$ is ϵ'-testable with $s = s(\epsilon') \in \mathbb{N}$ samples and $q = q(\epsilon') \in \mathbb{N}$ queries for every $\epsilon' \in (0,1)$, and $n \geq 2^{\mathrm{poly}(2^q, s, 1/\epsilon')}$, then for every $\epsilon \in (0,1)$ there exists an algorithm that given access to an unknown distribution μ over $\{0,1\}^n$, performs at most $\exp\left(\exp\left(2^{q(\Omega(\epsilon))} \cdot \mathrm{poly}\left(s(\Omega(\epsilon)), q(\Omega(\epsilon)), 1/\epsilon\right)\right)\right)$ queries, and outputs a value $\widetilde{\boldsymbol{d}}$ such that with probability at least $2/3$ it holds that $|\widetilde{\boldsymbol{d}} - d_{\mathrm{EM}}(\mu, \mathcal{P})| \leq \epsilon$.*

Theorem 2 follows directly from the following lemma about the existence of an (ϵ_1, ϵ_2)-tolerant test for any index-invariant property $\mathcal{P}$ which admits an ϵ-test for $\epsilon = (\epsilon_2 - \epsilon_1)/12$ with s samples and q queries (see Claim 2 in [PRR06]).

Lemma 6.1. *Suppose that an index-invariant property of distributions $\mathcal{P}$ has an ϵ-test with $s(\epsilon)$ samples and $q(\epsilon)$ queries for every $\epsilon \in (0,1)$. Then, for every $0 < \epsilon_1 < \epsilon_2 < 1$, Tolerant-Tester (see Figure 7) is an (ϵ_1, ϵ_2)-tolerant tester for $\mathcal{P}$ that makes at most*

$$\exp\left(\exp\left(2^q \cdot \mathrm{poly}\left(s((\epsilon_2 - \epsilon_1)/12), q((\epsilon_2 - \epsilon_1)/12), 1/(\epsilon_2 - \epsilon_1)\right)\right)\right)$$

queries to the samples obtained from the unknown input distribution.

The proof of Lemma 6.1 follows from the following two lemmas, proved in Section 6.1 and Section 6.2, respectively.

Lemma 6.2. *Fix $0 < \epsilon_1 < \epsilon_2 < 1$, and set $s = s((\epsilon_2 - \epsilon_1)/12)$ and $q = q((\epsilon_2 - \epsilon_1)/12)$. For any $n \geq 18q^2(s+1)/(\epsilon_2 - \epsilon_1)$, if a distribution μ over $\{0,1\}^n$ is ϵ_1-close to $\mathcal{P}$, then Tolerant-Tester accepts it with probability at least $2/3$.*

Lemma 6.3. *Fix $0 < \epsilon_1 < \epsilon_2 < 1$, and set $s = s((\epsilon_2 - \epsilon_1)/12)$ and $q = q((\epsilon_2 - \epsilon_1)/12)$. For any $n \geq 2^{\mathrm{poly}(2^q, s, 1/(\epsilon_2 - \epsilon_1))}$, if a distribution μ over $\{0,1\}^n$ is ϵ_2-far from $\mathcal{P}$, then Tolerant-Tester rejects it with probability at least $2/3$.*

Proof of Lemma 6.1: Assuming Lemma 6.2 and Lemma 6.3 hold, the correctness of Tolerant-Tester is immediate. To bound the query complexity, note that the algorithm performs all of its queries in steps (2) and (3). Thus, by the choice of parameters, Lemma 5.11 and Lemma 5.12, the size of U is at most $2^q \cdot \mathrm{poly}(s, q, 1/(\epsilon_2 - \epsilon_1))$, and the query complexity of the algorithm is at most $\exp\left(\exp\left(2^q \cdot \mathrm{poly}\left(s, q, 1/(\epsilon_2 - \epsilon_1)\right)\right)\right)$. ∎

44

Procedure Tolerant-Tester$(\mu, \mathcal{P}, \epsilon_1, \epsilon_2, s, q)$

Input: Sample and query access to μ, index-invariant property $\mathcal{P}$, and parameters $\epsilon_1, \epsilon_2 \in (0,1)$ such that $\epsilon_1 < \epsilon_2$, number of samples s and queries q.
Output: Accept if μ is ϵ_1-close to $\mathcal{P}$, and **Reject** if μ is ϵ_2-far from $\mathcal{P}$.

1. Let $\delta = \frac{(\epsilon_2 - \epsilon_1)^2}{10^5 \cdot q^3 (s+1)^6}$, $k = q - 1$, $k' = \Theta(2^{8k/\delta}/\delta^{38})$ and $\rho = \left(\left\lceil \frac{36sq}{\epsilon_2 - \epsilon_1} \right\rceil \right)^{-1}$.

2. Call Find-Weakly-Robust-Detailing$(\delta, k', 1/6)$ on μ and obtain the detailing μ^U with respect to a set of variables $U \subseteq [n]$.

3. Call Estimate-Parameters$(\mu, U, \rho, 1/6)$ to obtain $(\tilde{\boldsymbol{\eta}}, \tilde{\boldsymbol{\Lambda}})$ where $\tilde{\boldsymbol{\eta}}$ is the weight distribution over 2^U and $\tilde{\boldsymbol{\Lambda}}$ is the type distribution over $[0,1]^{2^U}$.

4. For each B of size $|B| \leq 2^{k/\delta}$, any $\frac{2\rho}{2^U \cdot |B|}$-quantized detailing η' of $\tilde{\boldsymbol{\eta}}$ with respect to B, and any $\frac{\rho}{(1/\rho + 1)^{2^{|U|} \cdot |B|}}$-quantized type distribution $\tilde{\Upsilon}$ on $\{0, \rho, 2\rho, \ldots, 1\}^{2^U \times B}$, do the following:

   (a) Call Accept-Probability$(s, q, \tilde{\Upsilon}, \eta', \frac{\epsilon_2 - \epsilon_1}{12})$. If the output is at least $1/2$, go to the next step. Otherwise, go to the next option for B, η' and $\tilde{\Upsilon}$.

   (b) Compute $d_{\mathrm{EM}}^{\eta'}(\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \tilde{\Upsilon})$, where $\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}$ is the flat extension of $\tilde{\boldsymbol{\Lambda}}$ with respect to B. If the computed distance is at most $\frac{\epsilon_1 + \epsilon_2}{2}$, output **Accept** and terminate.

5. Output **Reject**.

Figure 7: Description of the Tolerant-Tester procedure.

6.1 Completeness (Proof of Lemma 6.2)

In this section, we prove that if the unknown distribution μ is ϵ_1-close to the index-invariant property $\mathcal{P}$, then it will be accepted by Tolerant-Tester (Figure 7) with probability at least $2/3$.

Lemma 6.4. *Fix $\delta \in (0,1)$, $\ell \in \mathbb{N}$, let μ be a distribution over $\{0,1\}^n$ and let ξ be a (δ, ℓ)-robust detailing of μ with respect to some set A. In addition, let ξ' be a refinement of ξ with respect to a set B of size ℓ, and let $\xi_{\langle \eta \rangle} \stackrel{\text{def}}{=} \xi_{\langle \xi'|_{2,3} \rangle}$ be a flat refinement of ξ with respect to $\eta \stackrel{\text{def}}{=} \xi'|_{2,3}$. Then, letting Λ' and $\Lambda_{\langle B \rangle}$ be the type distributions of ξ' and $\xi_{\langle \eta \rangle}$ respectively, we have that $d_{\mathrm{EM}}^{\eta}(\Lambda', \Lambda_{\langle B \rangle}) \leq \sqrt{\delta}$.*

Proof: By definition of the index and the fact that ξ is (δ, ℓ)-robust, we can say the following:

$$\delta \geq \mathrm{Ind}(\xi') - \mathrm{Ind}(\xi) = \mathop{\mathbf{E}}_{i \sim [n]}\left[\mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \eta}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1]^2 - \mathop{\mathbf{E}}_{\boldsymbol{a} \sim \eta|_1}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}}[\boldsymbol{x}_i = 1]^2\right]\right]\right]$$

$$= \mathop{\mathbf{E}}_{i \sim [n]}\left[\mathop{\mathbf{E}}_{\boldsymbol{a} \sim \eta|_1}\left[\mathop{\mathbf{E}}_{\boldsymbol{b} \sim \eta|_2^{1:\boldsymbol{a}}}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1]^2\right] - \mathop{\mathbf{E}}_{\boldsymbol{b} \sim \eta|_2^{1:\boldsymbol{a}}}\left[\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1]\right]^2\right]\right]$$

$$= \mathop{\mathbf{E}}_{i \sim [n]}\left[\mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \eta}\left[\left(\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}}[\boldsymbol{x}_i = 1]\right)^2\right]\right]. \tag{10}$$

Now, letting Λ be the type distribution of ξ and using the fact that $\xi_{\langle \eta \rangle}$ is a flat refinement of ξ, by Observation 3.19 (using the η-weighted ℓ_1 distance as the metric):

$$d_{\mathrm{EM}}^\eta(\Lambda', \Lambda_{\langle \eta \rangle}) \leq \mathop{\mathbf{E}}_{i \sim [n]} \mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \eta}\left[\left|\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi_{\langle \eta \rangle}|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1]\right|\right]$$

$$\leq \mathop{\mathbf{E}}_{i \sim [n]} \mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \eta}\left[\left|\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}}[\boldsymbol{x}_i = 1]\right| - \left|\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi_{\langle \eta \rangle}|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1]\right|\right]$$

$$= \mathop{\mathbf{E}}_{i \sim [n]} \mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \eta}\left[\left|\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}}[\boldsymbol{x}_i = 1]\right|\right]$$

$$\leq \sqrt{\mathop{\mathbf{E}}_{i \sim [n]} \mathop{\mathbf{E}}_{(\boldsymbol{a},\boldsymbol{b}) \sim \eta}\left[\left(\mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi'|_1^{2,3:(\boldsymbol{a},\boldsymbol{b})}}[\boldsymbol{x}_i = 1] - \mathop{\mathbf{Pr}}_{\boldsymbol{x} \sim \xi|_1^{2:\boldsymbol{a}}}[\boldsymbol{x}_i = 1]\right)^2\right]} \leq \sqrt{\delta},$$

where in the last line we used the Cauchy-Schwartz inequality (Lemma 2.34), followed by Equation (10). ∎

In the following we will deal with a detailing Ξ of a transfer distribution between two distributions μ and τ over $\{0,1\}^n$ with respect to a set A, rather than with a detailing of a "single distribution" over $\{0,1\}^n$. So typically such Ξ can be viewed as a distribution over $\{0,1\}^n \times \{0,1\}^n \times A$, and in particular $\Xi|_{1,3}$ would be a detailing of μ and $\Xi|_{2,3}$ would be a detailing of τ. The following lemma connects the EMD between μ and τ to a distance between the type distributions of the two detailings resulting from such Ξ.

Lemma 6.5. *Let T be a transfer function realizing the EMD between μ and τ, and let Ξ be a detailing of T with respect to A. Additionally, let Λ^* be the type distribution of the detailing $\Xi|_{1,3}$ over μ, let Υ^* be the type distribution of the detailing $\Xi|_{2,3}$ of τ, and set $\eta = \Xi|_3$ to be the weight distribution (common to all detailings mentioned here). Then, $d_{\mathrm{EM}}^\eta(\Lambda^*, \Upsilon^*) \leq d_{\mathrm{EM}}(\mu, \tau)$.*

Proof: We first note that by definition of the EMD:

$$d_{\mathrm{EM}}(\mu, \tau) = \underset{(\boldsymbol{x},\boldsymbol{y})\sim T}{\mathbf{E}}\left[d_H(\boldsymbol{x},\boldsymbol{y})\right] = \underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left[\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{E}}\left[d_H(\boldsymbol{x},\boldsymbol{y})\right]\right]$$

$$= \underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left[\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{E}}\left[\underset{i\sim[n]}{\mathbf{E}}\left[\mathbf{1}_{\{\boldsymbol{x}_i\neq\boldsymbol{y}_i\}}\right]\right]\right]$$

$$= \underset{i\sim[n]}{\mathbf{E}}\left[\underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left[\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{E}}\left[\mathbf{1}_{\{\boldsymbol{x}_i\neq\boldsymbol{y}_i\}}\right]\right]\right]$$

$$= \underset{i\sim[n]}{\mathbf{E}}\left[\underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left[\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=0,\boldsymbol{y}_i=1] + \underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=1,\boldsymbol{y}_i=0]\right]\right]. \qquad (11)$$

On the other hand, by Observation 3.19, we can say that:

$$d_{\mathrm{EM}}^{\eta}(\Lambda^*, \Upsilon^*) \leq \underset{i\sim[n]}{\mathbf{E}}\left[\underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left|\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=1] - \underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{y}_i=1]\right|\right]$$

$$= \underset{i\sim[n]}{\mathbf{E}}\left[\underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left|\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=1,\boldsymbol{y}_i=0] - \underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=0,\boldsymbol{y}_i=1]\right|\right]$$

$$\leq \underset{i\sim[n]}{\mathbf{E}}\left[\underset{\boldsymbol{a}\sim\Xi|_3}{\mathbf{E}}\left[\underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=1,\boldsymbol{y}_i=0] + \underset{(\boldsymbol{x},\boldsymbol{y})\sim\Xi|^{3:\boldsymbol{a}}}{\mathbf{Pr}}[\boldsymbol{x}_i=0,\boldsymbol{y}_i=1]\right]\right].$$

Combined with Equation (11) the lemma follows. ∎

As a heads-up, the above lemma will be eventually used for a detailing Ξ with respect to a cross-product set $A \times B$, because it will be constructed from a join (see Definition 2.4) involving two detailings of two respective individual distributions. We are now ready to prove the completeness of Tolerant-Tester.

Proof of Lemma 6.2: Suppose that there exists a distribution $\tau \in \mathcal{P}$ for which $d_{\mathrm{EM}}(\mu,\tau) \leq \epsilon_1$, and let T be a transfer function exhibiting the distance. We will show that in such case, the algorithm Tolerant-Tester accepts. Let $k = q-1$, $\delta = \frac{(\epsilon_2-\epsilon_1)^2}{10^5 \cdot (s+1)^6 q^3}$ and $k' = \Theta\left(\frac{2^{8k/\delta}}{\delta^{38}}\right)$. Let μ^U be the detailing returned by the call to Find-Weakly-Robust-Detailing($\delta, k', 1/6$), let $\eta = \mu^U|_2$ be its weight distribution and let Λ be its type distribution.

Note that by the guarantees of Find-Weakly-Robust-Detailing (Lemma 5.11) we have that with probability at least $5/6$, the returned set U defines a (δ, k')-weakly robust detailing. In addition, from the guarantees of Estimate-Parameters (Lemma 5.12) we have that $d_{\mathrm{TV}}(\tilde{\boldsymbol{\eta}}, \eta) \leq \rho$ and $d_{\mathrm{EM}}^{\eta}(\tilde{\boldsymbol{\Lambda}}, \Lambda) \leq \rho$ with probability at least $5/6$. We henceforth condition on the intersection of the above events (which happens with probability at least $2/3$), and prove that in this case the algorithm will indeed accept.

Set $\Xi' = T \bowtie \mu^U$ (see Definition 2.4, where in the join operation unify the first coordinate of T with the first coordinate of μ^U), considering it as a distribution on $\{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{|U|}$, and note that this is a detailing of T with respect to $A \overset{\text{def}}{=} \{0,1\}^{|U|}$. Also set $\zeta = \Xi'|_{2,3}$, and note that this is a detailing of τ with respect to A. However, ζ is a general detailing of τ, not necessarily one defined by variables.

Next, use Lemma 3.23 to find a refinement ζ^V which is (δ, k)-weakly robust, where $V \subset [n]$ is a variable set of size at most k/δ. By Lemma 3.26, the detailing ζ^V is also $\left(\frac{\epsilon_2-\epsilon_1}{100(s+1)}, q\right)$-good.

We now set $\Xi = \Xi' \bowtie \zeta^V$. We consider it to be a distribution over $\{0,1\}^n \times \{0,1\}^n \times (A \times B)$ for

47

$B \stackrel{\text{def}}{=} \{0,1\}^{|V|}$, noting that it is a refinement of Ξ' with respect to B. We also note that $\Xi|_{2,3} = \zeta^V$, and that $\Xi|_{1,3}$ is a refinement of ξ with respect to B (and is in particular a detailing of μ, but no longer one defined by variables). We now let $\eta^* = \Xi|_3$ be the weight distribution of Ξ (this is a distribution over $A \times B$ which is a detailing of η with respect to B), let Λ^* be the type distribution of $\Xi|_{1,3}$, and let Υ^* be the type distribution of $\Xi|_{2,3}$.

By Lemma 2.23 and Lemma 2.28, there exists a $\frac{2\rho}{|A| \cdot |B|}$-quantized detailing (in the sense of Definition 3.3) $\widehat{\eta}$ of $\xi|_2 = \eta|_1$ with respect to B and a $\frac{\rho}{(1/\rho+1)^{|A| \cdot |B|}}$-quantized distribution $\tilde{\Upsilon}$ over $\{0, \rho, \ldots, 1\}^{A \times B}$, for which $d_{\text{TV}}(\eta', \eta^*) \leq \rho$ and $d_{\text{EM}}^{\eta^*}(\tilde{\Upsilon}, \Upsilon^*) \leq \rho$.

Now note that $\widetilde{\boldsymbol{\eta}}(v) = 0$ whenever $\widehat{\eta}(v) = \xi|_2(v) = 0$, and let $\eta' = \widetilde{\boldsymbol{\eta}} \rhd \widehat{\eta}$ be the adjustment of $\widehat{\eta}$ to $\widetilde{\boldsymbol{\eta}}$ (see Definition 2.5). While $\widehat{\eta}$ may not be a distribution that is considered by the loop in Step 4 of Tolerant-Tester, the distribution η' (as a ρ-quantized detailing of $\widetilde{\boldsymbol{\eta}}$) is considered there. Note that by Lemma 2.11 $d_{\text{TV}}(\widehat{\eta}, \eta') \leq \rho$, and hence by the triangle inequality $d_{\text{TV}}(\eta', \eta^*) \leq 2\rho$. To conclude the completeness proof, we will show that the algorithm will accept μ through the consideration of the pair $(\eta', \tilde{\Upsilon})$.

Using Lemma 4.5 we have that Accept-Probability$(s, q, \tilde{\Upsilon}, \eta', \frac{\epsilon_2 - \epsilon_1}{12})$ deviates from the true acceptance probability by at most $1/20$. Therefore, since the canonical tester accepts τ with probability at least $2/3$, Accept-Probability returns a value which is at least $2/3 - 1/20 \geq 1/2$, so the pair $(\eta', \tilde{\Upsilon})$ passes Step (4a) in Tolerant-Tester.

It remains to bound $d_{\text{EM}}^{\eta'}(\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \tilde{\Upsilon})$. Note that using Lemma 3.32 and the fact that μ^U is (δ, k')-weakly robust, we have that μ^U is also $(2\delta, 2^{k/\delta})$-robust. Therefore, letting $\Lambda_{\langle B \rangle}$ (respectively $\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}$) be the flat extension of Λ (respectively $\tilde{\boldsymbol{\Lambda}}$) over B, from Lemma 6.4 we have that $d_{\text{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \Lambda^*) \leq \sqrt{2\delta}$. In addition, since taking a flat extension does not change the distance, $d_{\text{EM}}^{\eta^*}(\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \Lambda_{\langle B \rangle}) = d_{\text{EM}}^{\eta}(\tilde{\boldsymbol{\Lambda}}, \Lambda) \leq \frac{\epsilon_2 - \epsilon_1}{12}$. Using Lemma 2.29, Lemma 6.5 and the triangle inequality:

$$
\begin{aligned}
d_{\text{EM}}^{\eta'}(\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \tilde{\Upsilon}) &\leq d_{\text{EM}}^{\eta^*}(\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \tilde{\Upsilon}) + 2\rho \\
&\leq d_{\text{EM}}^{\eta^*}(\tilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \Lambda_{\langle B \rangle}) + d_{\text{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \Lambda^*) + d_{\text{EM}}^{\eta^*}(\Lambda^*, \Upsilon^*) + d_{\text{EM}}^{\eta^*}(\Upsilon^*, \tilde{\Upsilon}) + 2\rho \\
&\leq \frac{\epsilon_2 - \epsilon_1}{12} + \sqrt{2\delta} + \epsilon_1 + 3\rho \\
&\leq \frac{\epsilon_2 - \epsilon_1}{12} + \frac{\epsilon_2 - \epsilon_1}{5^{5/2} s^3 q^{3/2}} + \epsilon_1 + \frac{\epsilon_2 - \epsilon_1}{9sq} \\
&\leq \frac{\epsilon_2 + \epsilon_1}{2}.
\end{aligned}
$$

Therefore, we have that (conditioned on the probability $2/3$ event described at the beginning of the proof), the algorithm Tolerant-Tester indeed accepts through $(\eta', \tilde{\Upsilon})$, as required. ∎

6.2 Soundness (Proof of Lemma 6.3)

Next, we will show that Tolerant-Tester is sound. That is, if μ is ϵ_2-far from the index-invariant property $\mathcal{P}$, then it will be rejected by Tolerant-Tester with probability at least $2/3$.

We define a procedure (used later only for an existence proof) that given a detailing ξ over A of a distribution μ, a detailing η of $\xi|_2$ with respect to B, a "target" type distribution Λ_t on $[0,1]^{A \times B}$, and an implementation $H : [n] \to ([0,1]^{A \times B})^2$ of a transfer function between the "source" type

distribution $\Lambda_s = \Lambda_{\langle B \rangle}$ of $\xi_{\langle \eta \rangle}$ and Λ_t which extends the implementation of Λ_s that is demonstrated by $\xi_{\langle \eta \rangle}$, produces a sample from a detailing Ξ of a transfer distribution between μ and some τ, so that the corresponding detailing of τ with respect to $A \times B$ has type distribution Λ_t, and the distance between μ and τ is bounded by the average of the distances between the types provided in H.

Procedure Change-types(ξ, η, H)

Input: A detailing ξ of μ with respect to A, a detailing η of $\xi|_2$ with respect to B, and an implementation $H : [n] \to ([0,1]^{A \times B})^2$ of a transfer distribution between the type distribution $\Lambda_s = \Lambda_{\langle B \rangle}$ of $\xi_{\langle \eta \rangle}$ and a target type distribution Λ_t, such that H extends the implementation of Λ_s demonstrated by $\xi_{\langle \eta \rangle}$.

Output: A sample from a distribution Ξ over $\{0,1\}^n \times \{0,1\}^n \times (A \times B)$ which is a detailing of a transfer distribution between μ and some τ.

1. Draw $(\boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b}) \sim \xi_{\langle \eta \rangle}$.

2. For every $i \in [n]$ independently, set $\boldsymbol{y}_i$ conditioned on the value of $\boldsymbol{x}_i$. If $\boldsymbol{x}_i = 1$, then set $\boldsymbol{y}_i = 0$ with probability $\max\left\{0, \frac{(H(i))_1(\boldsymbol{a},\boldsymbol{b}) - (H(i))_2(\boldsymbol{a},\boldsymbol{b})}{(H(i))_1(\boldsymbol{a},\boldsymbol{b})}\right\}$, and if $\boldsymbol{x}_i = 0$, then set $\boldsymbol{y}_i = 1$ with probability $\max\left\{0, \frac{(H(i))_2(\boldsymbol{a},\boldsymbol{b}) - (H(i))_1(\boldsymbol{a},\boldsymbol{b})}{1 - (H(i))_1(\boldsymbol{a},\boldsymbol{b})}\right\}$.

3. return $(\boldsymbol{x}, \boldsymbol{y}, (\boldsymbol{a}, \boldsymbol{b}))$.
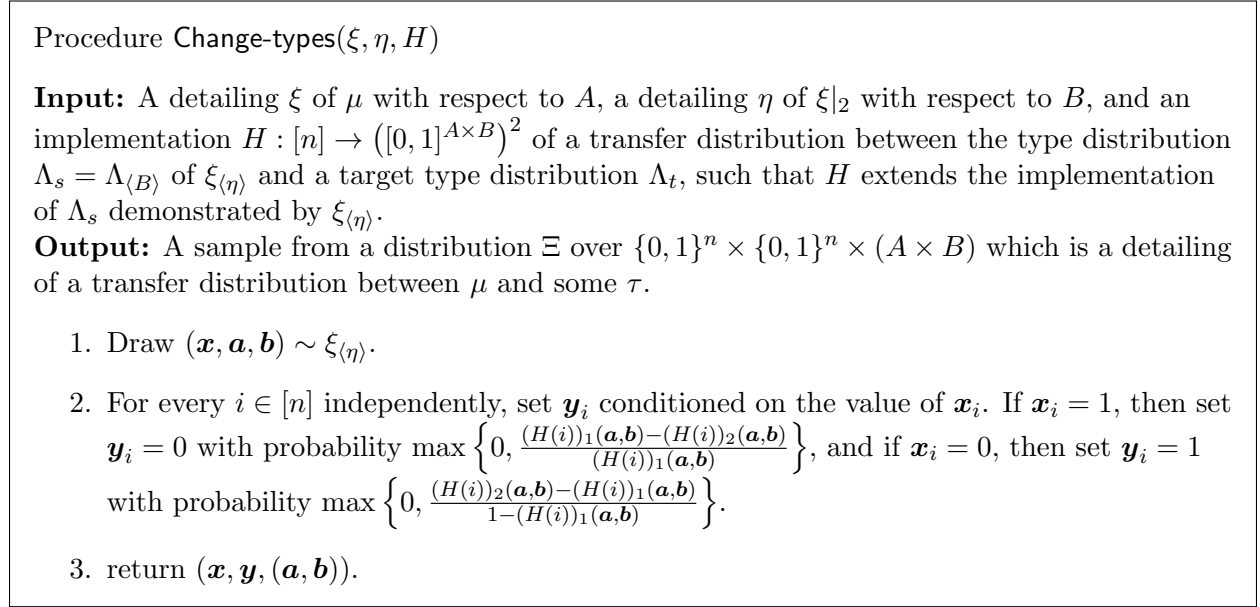

Figure 8: A description of the Change-types procedure

We will later use Change-types to show that if Tolerant-Tester does not reject with high probability then there exists some distribution τ which is close to μ and is not rejectable by a $\frac{\epsilon_2 - \epsilon_1}{12}$-test for $\mathcal{P}$, meaning that μ itself is not very far from $\mathcal{P}$. But first we need to prove its properties, starting with the following trivial observation (which follows from the construction of Ξ).

Observation 6.6. *The distribution Ξ of the output of Change-types(ξ, η, H) satisfies $\Xi|_{1,3} = \xi_{\langle \eta \rangle}$, and in particular the type distribution of $\Xi|_{1,3}$ equals $\Lambda_{\langle B \rangle}$.*

The following shows that the detailing $\Xi|_{2,3}$ of the distribution $\tau = \Xi|_2$ indeed has the required target type distribution.

Lemma 6.7. *The procedure Change-types produces a sample from a distribution Ξ such that the detailing $\Xi|_{2,3}$ of τ admits the type distribution Λ_t.*

Proof: We prove for every i that indeed $\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y}) \sim \Xi|^{3:(a,b)}}[\boldsymbol{y}_i = 1] = (H(i))_2(a,b)$, which means that the corresponding type for this coordinate is indeed $(H(i))_2$.

$$\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y}) \sim \Xi|^{3:(a,b)}}[\boldsymbol{y}_i = 1] = \mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})}[\boldsymbol{y}_i = 1 \mid \boldsymbol{x}_i = 0] \cdot \mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})}[\boldsymbol{x}_i = 0] + \mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})}[\boldsymbol{y}_i = 1 \mid \boldsymbol{x}_i = 1] \cdot \mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y})}[\boldsymbol{x}_i = 1].$$

In the case $(H(i))_1(a,b) \geq (H(i))_2(a,b)$, this gives

$$\mathbf{Pr}_{(\boldsymbol{x},\boldsymbol{y}) \sim \Xi|^{3:(a,b)}}[\boldsymbol{y}_i = 1] = 0 + (H(i))_1(a,b) \cdot \left(1 - \frac{(H(i))_1(a,b) - (H(i))_2(a,b)}{(H(i))_1(a,b)}\right) = (H(i))_2(a,b).$$

In the case $(H(i))_1(a, b) < (H(i))_2(a, b)$, this gives

$$(1 - (H(i))_1(a, b)) \cdot \frac{(H(i))_2(a, b) - (H(i))_1(a, b)}{1 - (H(i))_1(a, b)} + (H(i))_1(a, b) = (H(i))_2(a, b).$$

■

The next lemma bounds the distance between $\mu = \Xi|_1$ and $\tau = \Xi|_2$.

Lemma 6.8. *The procedure* Change-types *produces a distribution Ξ such that*

$$d_{\mathrm{EM}}(\Xi|_1, \Xi|_2) \leq \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim \eta} \left[\left| (H(i))_2(\boldsymbol{a}, \boldsymbol{b}) - (H(i))_1(\boldsymbol{a}, \boldsymbol{b}) \right| \right] \right].$$

Proof: Noting that $\Xi|_3 = \eta$, we have that:

$$d_{\mathrm{EM}}(\Xi|_1, \Xi|_2) \leq \mathop{\mathbf{E}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{1,2}} [d_H(\boldsymbol{x}, \boldsymbol{y})] = \mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim \eta} \left[\mathop{\mathbf{E}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [d_H(\boldsymbol{x}, \boldsymbol{y})] \right]$$

$$= \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim \eta} \left[\mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 0 \wedge \boldsymbol{y}_i = 1] + \mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 1 \wedge \boldsymbol{y}_i = 0] \right] \right].$$

(12)

For a fixed $i \in [n]$ and $(a, b) \in A \times B$ we have that

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 0 \wedge \boldsymbol{y}_i = 1] = \mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{y}_i = 1 \mid \boldsymbol{x}_i = 0] \cdot \mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 0]$$

$$= (1 - (H(i))_1(a, b)) \cdot \max \left\{ 0, \frac{(H(i))_2(a, b) - (H(i))_1(a, b)}{1 - (H(i))_1(a, b)} \right\}$$

$$= \max \left\{ 0, (H(i))_2(a, b) - (H(i))_1(a, b) \right\},$$

and similarly,

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 1 \wedge \boldsymbol{y}_i = 0] = \max \left\{ 0, (H(i))_1(a, b) - (H(i))_2(a, b) \right\}.$$

Therefore,

$$\mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 0 \wedge \boldsymbol{y}_i = 1] + \mathop{\mathbf{Pr}}_{(\boldsymbol{x}, \boldsymbol{y}) \sim \Xi|_{3:(\boldsymbol{a}, \boldsymbol{b})}} [\boldsymbol{x}_i = 1 \wedge \boldsymbol{y}_i = 0] = \left| (H(i))_2(a, b) - (H(i))_1(a, b) \right|.$$

Plugging into Inequality (12), we have the final result:

$$d_{\mathrm{EM}}(\Xi|_1, \Xi|_2) \leq \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim \eta} \left[\left| (H(i))_2(\boldsymbol{a}, \boldsymbol{b}) - (H(i))_1(\boldsymbol{a}, \boldsymbol{b}) \right| \right] \right].$$

■

The final and rather important property of the output distribution of Change-types is the preservation of goodness. Namely, that if ξ was an (ϵ, q)-good detailing of the unknown input distribution μ, then $\Xi|_{2,3}$ will be an (ϵ, q)-good detailing of the "new" τ.

Lemma 6.9. *Fix $\epsilon \in (0, 1)$ and $q \in \mathbb{N}$. Let ξ be a detailing of μ with respect to A, and let η be a distribution over $A \times B$ which is a detailing of $\xi|_2$ with respect to B. If ξ is (ϵ, q)-good, then the detailing $\Xi|_{2,3}$ of $\tau = \Xi|_2$, as obtained by the output distribution of* Change-types, *is also (ϵ, q)-good.*

50

Proof: Since ξ is (ϵ, q)-good, there exists $J \subseteq A$ with $\mathbf{Pr}_{\xi|_2}[J] > 1 - \epsilon$ such that for any $a \in J$, at least $(1 - \epsilon)n^q$ of the q-tuples in $[n]^q$ are ϵ-independent. Consider the flat extension $\xi_{\langle \eta \rangle}$ with respect to η. Since $\xi_{\langle \eta \rangle}$ is a flat extension, for any $a \in J$ and $\alpha \in [n]^q$ which is ϵ-independent with respect to $\xi|_1^{2:a}$, the tuple α is also ϵ-independent with respect to $\xi_{\langle \eta \rangle}|_1^{2:(a,b)} = \xi|^{2:a}$ for all $b \in B$.

Next we claim that if a tuple $\alpha \in [n]^q$ is ϵ-independent with respect to $\xi_{\langle \eta \rangle}|_1^{2:(a,b)}$, then it is also ϵ-independent with respect to $\nu \stackrel{\text{def}}{=} \Xi|_2^{3:(a,b)}$. Let $(j_1, \ldots, j_q) \in [n]^q$ be an ϵ-independent tuple with respect to $\xi_{\langle \eta \rangle}|_1^{2:(a,b)}$. Then, we have that

$$
d_{\mathrm{TV}}\left(\left(\xi_{\langle \eta \rangle}|_1^{2:(a,b)} \right)|_{\{j_1, \ldots, j_q\}}, \prod_{\ell \in [q]} \left(\xi_{\langle \eta \rangle}|_1^{2:(a,b)} \right)|_{j_\ell} \right) \leq \epsilon.
$$

Let σ denote an optimal coupling from the distribution $\left(\xi_{\langle \eta \rangle}|_1^{2:(a,b)} \right)|_{\{j_1, \ldots, j_q\}}$ to the corresponding product distribution $\prod_{\ell \in [q]} \left(\xi_{\langle \eta \rangle}|_1^{2:(a,b)} \right)|_{j_\ell}$, in the sense that

$$
d_{\mathrm{TV}}\left(\left(\xi_{\langle \eta \rangle}|_1^{2:(a,b)} \right)|_{\{j_1, \ldots, j_q\}}, \prod_{\ell \in [q]} \left(\xi_{\langle \eta \rangle}|_1^{2:(a,b)} \right)|_{j_\ell} \right) = \mathop{\mathbf{E}}_{(\boldsymbol{x}, \boldsymbol{x}') \sim \sigma} \left[\mathbf{1}_{\{\boldsymbol{x} \neq \boldsymbol{x}'\}} \right]
$$

We construct a coupling σ' from the distribution $\nu|_{\{j_1, \ldots, j_q\}}$ to the corresponding product distribution $\prod_{\ell \in [q]} \nu|_{j_\ell}$ and and prove that

$$
\mathop{\mathbf{E}}_{(\boldsymbol{y}, \boldsymbol{y}') \sim \sigma'} \left[\mathbf{1}_{\{\boldsymbol{y} \neq \boldsymbol{y}'\}} \right] \leq \mathop{\mathbf{E}}_{(\boldsymbol{x}, \boldsymbol{x}') \sim \sigma} \left[\mathbf{1}_{\{\boldsymbol{x} \neq \boldsymbol{x}'\}} \right].
$$

We define the coupling σ' as follows. We first sample $(\boldsymbol{x}, \boldsymbol{x}') \sim \sigma$. Then for every $\ell \in [q]$, if $\boldsymbol{x}_{j_\ell} = \boldsymbol{x}'_{j_\ell}$ then we apply Step (2) of Change-types for $\boldsymbol{x}_{j_\ell}$ to obtain $\boldsymbol{y}_{j_\ell}$ and set $\boldsymbol{y}'_{j_\ell} = \boldsymbol{y}_{j_\ell}$ (this means that we "use the same random coins" for handling $\boldsymbol{x}_{j_\ell}$ and $\boldsymbol{x}'_{j_\ell}$). If $\boldsymbol{x}_{j_\ell} \neq \boldsymbol{x}'_{j_\ell}$ then we just apply Step (2) of Change-types separately for $\boldsymbol{x}_{j_\ell}$ to obtain $\boldsymbol{y}_{j_\ell}$ and for $\boldsymbol{x}'_{j_\ell}$ to obtain $\boldsymbol{y}'_{j_\ell}$ (note that in this case only one of the two applications uses randomness). The last two things to note are that $\boldsymbol{y}$ has the same distribution as the output of Change-types over $\boldsymbol{x}$ restricted to $\{j_1, \ldots, j_q\}$, while $\boldsymbol{y}'$ distributes as the corresponding product, and that under this process $\boldsymbol{x} = \boldsymbol{x}'$ implies $\boldsymbol{y} = \boldsymbol{y}'$. Hence $\mathbf{E}_{(\boldsymbol{y}, \boldsymbol{y}') \sim \sigma'}[\mathbf{1}_{\{\boldsymbol{y} \neq \boldsymbol{y}'\}}] \leq \mathbf{E}_{(\boldsymbol{x}, \boldsymbol{x}') \sim \sigma}[\mathbf{1}_{\{\boldsymbol{x} \neq \boldsymbol{x}'\}}]$, as required.

Since the above argument holds for any ϵ-independent tuple with respect to $\xi_{\langle \eta \rangle}|_1^{2:(a,b)}$ for any $(a, b) \in A \times B$ (showing the tuple to be ϵ-independent also for $\Xi|_2^{3:(a,b)}$), it implies that (ϵ, q)-goodness is indeed transferred from $\xi_{\langle \eta \rangle}$ to $\Xi|_{2,3}$. ∎

Our application to this lemma would be to show that predictability still holds for the resulting detailing of τ (with respect to the weight distribution η and the type distribution Λ_t), just as it held for a good enough detailing ξ of μ.

Lemma 6.10. *Fix $\epsilon, \gamma \in (0, 1)$, $s, q \in \mathbb{N}$, a detailing ξ of μ with respect to A, and a distribution η over $A \times B$ extending $\xi|_2$. Let $\delta = \frac{\epsilon^6}{10^5 \cdot q^3 \cdot s^6}$, let $\mathsf{acc}_\gamma(\Xi|_2)$ denote the acceptance probability of the canonical tester with proximity γ applied on $\Xi|_2$, where Ξ is the output distribution of Change-types, and denote by Υ the type distribution of $\Xi|_{2,3}$. If ξ is $(\delta, q-1)$-weakly robust, then the output*

$\widetilde{acc}_\gamma(\Xi|_2)$ of the procedure Accept-Probability$(s, q, \eta, \Upsilon, \gamma)$ satisfies

$$|\widetilde{acc}_\gamma(\Xi|_2) - acc_\gamma(\Xi|_2)| \leq \epsilon/20.$$

Proof: Since ξ is (δ, ϵ)-weakly robust, by Lemma 3.26 it is also (ϵ', q)-good for some $\epsilon' \leq \frac{\epsilon}{1000s}$. By Lemma 6.9, we have that the detailing $\Xi|_{2,3}$ is also (ϵ', q)-good. Therefore, we can use Lemma 4.5 to conclude that $|\widetilde{acc}_\gamma(\Xi|_2) - acc_\gamma(\Xi|_2)| \leq \epsilon/20$, as required. ∎

Proof of Lemma 6.3: By the guarantees of Find-Weakly-Robust-Detailing (Lemma 5.11) we have that with probability at least $5/6$ the returned set U defines a (δ, k')-weakly robust detailing μ^U of μ. In addition, from the guarantees of Estimate-Parameters (Lemma 5.12) we have that $d_{\mathrm{TV}}(\tilde{\eta}, \eta) \leq \rho$ and $d_{\mathrm{EM}}^\eta(\tilde{\Lambda}, \Lambda) \leq \rho$ with probability at least $5/6$, where $\eta = \mu|_U$ is the weight distribution of μ^U and Λ is its type distribution. We henceforth condition on the intersection of the above events (which happens with probability at least $2/3$).

We will show that if the algorithm accepts, then there exists a distribution τ^*, which is accepted by the canonical test with probability larger than $1/3$ (and hence satisfies $d_{\mathrm{EM}}(\tau^*, \mathcal{P}) \leq \frac{\epsilon_2 - \epsilon_1}{12}$), satisfying also $d_{\mathrm{EM}}(\mu, \tau^*) \leq \frac{11\epsilon_2 + \epsilon_1}{12}$. By the triangle inequality, this will imply that $d_{\mathrm{EM}}(\mu, \mathcal{P}) \leq \epsilon_2$.

Let us denote $A = \{0, 1\}^{|U|}$, as in the proof of Lemma 6.2. Note that if Tolerant-Tester accepts, then there exist a $\frac{2\rho}{|A| \cdot |B|}$-quantized detailing η' of $\tilde{\eta}$ with respect to $A \times B$, and a $\frac{\rho}{(1 + 1/\rho)^{|A| \cdot |B|}}$-quantized type distribution $\tilde{\Upsilon}$ over $\{0, \rho, \ldots, 1\}^{A \times B}$, that satisfy $d_{\mathrm{EM}}^{\eta'}(\tilde{\Lambda}_{\langle B \rangle}, \tilde{\Upsilon}) \leq (\epsilon_1 + \epsilon_2)/2$ and such that Accept-Probability$(s, q, \tilde{\Upsilon}, \eta', \frac{\epsilon_2 - \epsilon_1}{12})$ returns a value that is at least $1/2$.

Noting that in particular $\tilde{\eta}(u) = 0$ whenever $\mu|_U(u) = 0$, let $\eta^* = \mu|_U \rhd \eta'$ be the adjustment of η' to $\mu|_U = \mu^U|_2$, and let $\mu^U_{\langle \eta^* \rangle}$ be the flat refinement of μ^U with respect to η^*. By Lemma 2.11 we have $d_{\mathrm{TV}}(\eta', \eta^*) \leq \rho$. Next we apply Lemma 2.28, assuming $n \geq \frac{(1/\rho + 1)^{2^{|U|} \cdot |B|}}{\rho} = 2^{\mathrm{poly}(2^q, s, 1/(\epsilon_2 - \epsilon_1))}$, to obtain a distribution Υ^* which is $\frac{1}{n}$-quantized such that $d_{\mathrm{TV}}(\Upsilon^*, \tilde{\Upsilon}) \leq \rho$.

By Lemma 2.29 we have $d_{\mathrm{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \Upsilon^*) \leq d_{\mathrm{EM}}^{\eta'}(\Lambda_{\langle B \rangle}, \Upsilon^*) + 2\rho$. Now let $h : [n] \to \{0, \rho, \ldots, 1\}^{A \times B}$ the implementation of $\Lambda_{\langle B \rangle}$ demonstrated by $\mu^U_{\langle \eta^* \rangle}$. We next note that we can apply Lemma 3.18 and obtain an implementation $H : [n] \to (\{0, \rho, \ldots, 1\}^{A \times B})^2$ of an optimal transfer function κ between $\Lambda_{\langle B \rangle}$ and Υ^* over η^*, with $(H(i))_1 = h(i)$ for all $i \in [n]$. Thus,

$$d_{\mathrm{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \Upsilon^*) = \mathop{\mathbf{E}}_{i \sim [n]}\left[\mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim \eta^*}\left[|(H(i))_2(\boldsymbol{a}, \boldsymbol{b}) - (H(i))_1(\boldsymbol{a}, \boldsymbol{b})| \right] \right]$$

We construct a distribution τ^* along with a detailing Ξ^* of a transfer distribution T^* from μ to τ^* with respect to $A \times B$ as follows. Recall that μ^U denotes the detailing of μ obtained in Step 2 of the algorithm.

We define Ξ^* as the distribution over the output of Change-types with respect to μ^U, η^* and H. That is, a sample $(\boldsymbol{x}, \boldsymbol{y}, (\boldsymbol{a}, \boldsymbol{b})) \sim \Xi^*$ is obtained by calling Change-types(μ^U, η^*, H). We will not take actual samples from Ξ^*, but only analyze the distance bounds that it implies.

Observe that by construction the distribution $\tau^* = \Xi^*|_2$ is supported on $\{0, 1\}^n$, and by Lemma 6.7 the detailing $\zeta = \Xi^*|_{2,3}$ of τ^* admits the target weight distribution η^* and type distribution Υ^*.

Then, by Lemma 6.8, using also the facts that $d_{\mathrm{EM}}^{\eta'}(\tilde{\Upsilon}, \Upsilon^*) \leq d_{\mathrm{TV}}(\tilde{\Upsilon}, \Upsilon^*)$ and $d_{\mathrm{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \tilde{\Lambda}_{\langle B \rangle}) =$

$d_{\mathrm{EM}}^{\mu|_U}(\Lambda, \widetilde{\boldsymbol{\Lambda}})$ we have that

$$
\begin{aligned}
d_{\mathrm{EM}}(\mu, \tau^*) &\leq \mathop{\mathbf{E}}_{i \sim [n]} \left[\mathop{\mathbf{E}}_{(\boldsymbol{a}, \boldsymbol{b}) \sim \eta^*} \left[|(H(i))_2(\boldsymbol{a}, \boldsymbol{b}) - (H(i))_1(\boldsymbol{a}, \boldsymbol{b})| \right] \right] = d_{\mathrm{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \Upsilon^*) \\
&\leq d_{\mathrm{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \widetilde{\boldsymbol{\Lambda}}_{\langle B \rangle}) + d_{\mathrm{EM}}^{\eta^*}(\widetilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \Upsilon^*) \\
&\leq d_{\mathrm{EM}}^{\eta^*}(\Lambda_{\langle B \rangle}, \widetilde{\boldsymbol{\Lambda}}_{\langle B \rangle}) + d_{\mathrm{EM}}^{\eta'}(\widetilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \Upsilon^*) + 2\rho \\
&\leq d_{\mathrm{EM}}^{\mu|_U}(\Lambda, \widetilde{\boldsymbol{\Lambda}}) + d_{\mathrm{EM}}^{\eta'}(\widetilde{\boldsymbol{\Lambda}}_{\langle B \rangle}, \tilde{\Upsilon}) + d_{\mathrm{TV}}(\tilde{\Upsilon}, \Upsilon^*) + 2\rho \\
&\leq \rho + \frac{\epsilon_1 + \epsilon_2}{2} + \rho + 2\rho \leq \frac{11\epsilon_2 + \epsilon_1}{12}
\end{aligned}
$$

It remains to show that τ^* is accepted by the canonical tester (with proximity parameter $\frac{\epsilon_2 - \epsilon_1}{12}$) with probability greater than $1/3$. Indeed, by our choice of parameters and using Lemma 6.10, we have that

$$
\left| \widetilde{\mathsf{acc}}_{\frac{\epsilon_2 - \epsilon_1}{12}}(\tau^*) - \mathsf{acc}_{\frac{\epsilon_2 - \epsilon_1}{12}}(\tau^*) \right| \leq 1/20,
$$

which implies that τ^* is accepted by the canonical tester with probability at least $1/2 - 1/20 > 1/3$, and the proof is complete. ∎

References

[ABEF17]   Noga Alon, Omri Ben-Eliezer, and Eldar Fischer. Testing hereditary properties of ordered graphs and matrices. In *Proceedings of the 58th Annual Symposium on Foundations of Computer Science (FOCS)*, 2017.

[ADK15]   Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

[AF24]   Tomer Adar and Eldar Fischer. Refining the adaptivity notion in the huge object model. In *Proceedings of the 28th International Conference on Randomization and Computation (RANDOM)*, 2024.

[AFKS00]   Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[AFL24]   Tomer Adar, Eldar Fischer, and Amit Levi. Support testing in the huge object model. In *Proceedings of the 28th International Conference on Randomization and Computation (RANDOM)*, 2024.

[AFNS06]   Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. In *Proceedings of the 38th annual ACM Symposium on Theory of Computing (STOC)*, 2006.

[BC17]   Tugkan Batu and Clément L. Canonne. Generalized uniformity testing. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2017.

[BCCZ18]   Christian Borgs, Jennifer T Chayes, Henry Cohn, and Yufei Zhao. An l_p theory of sparse graph convergence II: Ld convergence, quotients and right convergence. *The Annals of Probability*, 46(1):337–396, 2018.

[BCCZ19]   Christian Borgs, Jennifer Chayes, Henry Cohn, and Yufei Zhao. An l_p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 372(5):3019–3062, 2019.

[BCL$^+$06a]   Christian Borgs, Jennifer Chayes, László Lovász, Vera T Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proceedings of the 38th annual ACM Symposium on Theory of Computing (STOC)*, 2006.

[BCL$^+$06b]   Christian Borgs, Jennifer Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Counting graph homomorphisms. In *Topics in Discrete Mathematics: Dedicated to Jarik Nešetřil on the Occasion of his 60th Birthday*, pages 315–371. Springer, 2006.

[BCL$^+$08]   Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

[BCL$^+$12]   Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs II. multiway cuts and statistical physics. *Annals of Mathematics*, pages 151–219, 2012.

[BDKR05]   Tugkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005.

[BEFLY21]   Omri Ben-Eliezer, Eldar Fischer, Amit Levi, and Yuichi Yoshida. Limits of ordered graphs and their applications. In *Proceedings of the 12th Innovations in Theoretical Computer Science (ITCS)*, 2021.

[BFLR20]   Omri Ben-Eliezer, Eldar Fischer, Amit Levi, and Ron D. Rothblum. Hard properties with (very) short PCPPs and their applications. In *Proceedings of the 11th Innovations in Theoretical Computer Science Conference (ITCS)*, 2020.

[BS11]   Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Selected Works of Oded Schramm*, pages 533–545, 2011.

[BY22]   Arnab Bhattacharyya and Yuichi Yoshida. *Property Testing: Problems and Techniques*. Springer Nature, 2022.

[Can20]   Clément L Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing*, pages 1–100, 2020.

[Can22]   Clément L Canonne. Topics and techniques in distribution testing: A biased but representative sample. *Foundations and Trends® in Communications and Information Theory*, 19(6):1032–1198, 2022.

[CDVV14]   Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the 25th annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.

[CFG+22]   Sourav Chakraborty, Eldar Fischer, Arijit Ghosh, Gopinath Mishra, and Sayantan Sen. Exploring the gap between tolerant and non-tolerant distribution testing. In *Proceedings of the 26th International Conference on Randomization and Computation (RANDOM)*, pages 27:1–27:23, 2022.

[CFG+23]   Sourav Chakraborty, Eldar Fischer, Arijit Ghosh, Gopinath Mishra, and Sayantan Sen. Testing of index-invariant properties in the huge object model. In *The 36th Annual Conference on Learning Theory (COLT)*, 2023.

[CFSZ21]   David Conlon, Jacob Fox, Benny Sudakov, and Yufei Zhao. The regularity method for graphs with few 4-cycles. *Journal of the London Mathematical Society*, 104(5):2376–2401, 2021.

[CR20]   Leonardo Nagami Coregliano and Alexander A Razborov. Semantic limits of dense combinatorial objects. *Russian Mathematical Surveys*, 75(4):627, 2020.

[CS10]   Artur Czumaj and Christian Sohler. Sublinear-time algorithms. *Property testing: current research and surveys*, pages 41–64, 2010.

[CSY25]   Clément L. Canonne, Sayantan Sen, and Joy Qiping Yang. Settling the complexity of testing grainedness of distributions, and application to uniformity testing in the huge object model. In *Proceedings of the 16th Innovations in Theoretical Computer Science Conference (ITCS)*. To appear, 2025.

[DKN14]   Ilias Diakonikolas, Daniel M Kane, and Vladimir Nikishkin. Testing identity of structured distributions. In *Proceedings of the 26th annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014.

[DKS18]   Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Sharp bounds for generalized uniformity testing. In *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[DP09]   D.P. Dubhashi and A. Panconesi. *Concentration of Measure for the Analysis of Randomized Algoritms*. Cambridge University Press, 2009.

[FF06]   Eldar Fischer and Lance Fortnow. Tolerant versus intolerant testing for boolean properties. *Theory of Computing*, 2(9):173—-183, 2006.

[Fis04]   Eldar Fischer. The art of uninformed decisions: A primer to property testing. In *Current Trends in Theoretical Computer Science*, pages 229–263. World Scientific, 2004.

[FK99]   Alan Frieze and Ravi Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[FN07a]   Eldar Fischer and Ilan Newman. Testing of matrix-poset properties. *Combinatorica*, 27(3):293–327, 2007.

[FN07b]   Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM Journal on Computing*, 37(2):482–501, 2007.

[GGR98]   Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

[GHHS20] Frederik Garbe, Robert Hancock, Jan Hladký, and Maryam Sharifzadeh. Limits of latin squares. *arXiv preprint arXiv:2010.07854*, 2020.

[GKS23] Lior Gishboliner, Nick Kushnir, and Asaf Shapira. Testing versus estimation of graph properties, revisited. In *Proceedings of the 27th International Conference on Randomization and Computation (RANDOM)*, 2023.

[Gol17] Oded Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

[Gow97] William T Gowers. Lower bounds of tower type for szemerédi's uniformity lemma. *Geometric & Functional Analysis GAFA*, 7(2):322–337, 1997.

[GR11] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 68–75, 2011.

[GR23] Oded Goldreich and Dana Ron. Testing distributions of huge objects. *TheoretiCS*, 2, 2023.

[HKM+13] Carlos Hoppen, Yoshiharu Kohayakawa, Carlos Gustavo Moreira, Balázs Ráth, and Rudini Menezes Sampaio. Limits of permutation sequences. *Journal of Combinatorial Theory, Series B*, 103(1):93–113, 2013.

[Hoe94] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The collected works of Wassily Hoeffding*. Springer, 1994.

[Lov12] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Soc., 2012.

[LS06] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.

[LS07] László Lovász and Balázs Szegedy. Szemerédi's lemma for the analyst. *Geometric And Functional Analysis*, 17(1):252–270, 2007.

[LS10] László Lovász and Balázs Szegedy. Testing properties of graphs and functions. *Israel Journal of Mathematics*, 178(1):113–156, 2010.

[MS16] Guy Moshkovitz and Asaf Shapira. A short proof of gowers' lower bound for the regularity lemma. *Combinatorica*, 36(2):187–194, 2016.

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

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006.

[RD85] Vojtech Rödl and Richard A Duke. On graphs with small subgraphs of large chromatic number. *Graphs and Combinatorics*, 1(1):91–96, 1985.

[Ron08] Dana Ron. Property testing: A learning theory perspective. *Foundations and Trends® in Machine Learning*, 1(3):307–402, 2008.

[Ron10]   Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends® in Theoretical Computer Science*, 5(2):73–205, 2010.

[RS96]   Ronitt Rubinfeld and Madhu Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

[RS11]   Ronitt Rubinfeld and Asaf Shapira. Sublinear time algorithms. *SIAM Journal on Discrete Mathematics*, 25(4):1562–1588, 2011.

[Sim23]   Michael Simkin. The number of n-queens configurations. *Advances in Mathematics*, 427:109127, 2023.

[Sze75]   Endre Szemerédi. *Regular partitions of graphs.* Stanford University, 1975.

[Val11]   Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011.

[VV17]   Gregory Valiant and Paul Valiant. Estimating the unseen: improved estimators for entropy and other properties. *Journal of the ACM*, 64(6):37:1–37:41, 2017.