Property Testing of Joint Distributions using Conditional Samples

RISHIRAJ BHATTACHARYYA, NISER Bhubaneswar and HBNI, India
SOURAV CHAKRABORTY, Chennai Mathematical Institute and Indian Statistical Institute, Kolkata, India

In this paper, we consider the problem of testing properties of joint distributions under the Conditional Sampling framework. In the standard sampling model, sample complexity of testing properties of joint distributions are exponential in the dimension, resulting in inefficient algorithms for practical use. While recent results achieve efficient algorithms for product distributions with significantly smaller sample complexity, no efficient algorithm is expected when the marginals are not independent.

In this paper, we initialize the study of conditional sampling in the multidimensional setting. We propose a subcube conditional sampling model where the tester can condition on a (adaptively) chosen subcube of the domain. Due to its simplicity, this model is potentially implementable in many practical applications particularly when the distribution is a joint distribution over Σ^n for some set Σ.

We present algorithms for various fundamental properties of distributions, in the subcube-conditioning model and prove that the sample complexity is polynomial in the dimension n (and not exponential as in the traditional model). We present an algorithm for testing identity to a known distribution using $\tilde{O}(n^2)$-subcube-conditional samples, an algorithm for testing identity between two unknown distributions using $\tilde{O}(n^5)$-subcube-conditional samples and an algorithm for testing identity to a product distribution using $\tilde{O}(n^5)$-subcube-conditional samples.

The central concept of our technique involves an elegant chain rule which can be proved using basic techniques of probability theory, yet powerful enough to avoid the curse of dimensionality.

CCS Concepts: • **Theory of computation → Computational complexity and cryptography**;

Additional Key Words and Phrases: Distribution Testing, Conditional Sampling, Subcube Conditioning

ACM Reference Format:
Rishiraj Bhattacharyya and Sourav Chakraborty. 2018. Property Testing of Joint Distributions using Conditional Samples. *ACM Trans. Comput. Theory* 0, 0, Article 0 (2018), 21 pages.

1 INTRODUCTION

Property Testing of Distributions. The boom of Big Data Analytics has rejuvenated the well studied area of hypothesis testing over unknown distributions. In Computer Science, the study of this type of problems was initiated by Batu et al. [Batu *et al.* 2013] under the framework of "Property Testing" [Goldreich *et al.* 1998; Rubinfeld & Sudan 1996] In this framework, the "tester" draws independent samples from the distribution, and decides whether the distribution satisfies

Authors' addresses: Rishiraj Bhattacharyya, NISER Bhubaneswar and HBNI, School of Computer Sciences, Bhubaneswar, India, rishiraj.bhattacharyya@gmail.com; Sourav Chakraborty, Chennai Mathematical Institute and Indian Statistical Institute, Kolkata, Chennai, Kolkata, India, sourav@cmi.ac.in.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 Association for Computing Machinery.
1942-3454/2018/0-ART0 $15.00
https://doi.org/

a certain property $\mathcal{P}$ (null hypothesis) or is far from any distribution that satisfies $\mathcal{P}$ (alternate hypothesis).

Several properties of probability distributions have been studied in this framework. Testing whether the distribution is uniform [Batu *et al.* 2001; Chan *et al.* 2014; Goldreich & Ron 2011], testing identity between two unknown distributions (taking samples from both the distributions) [Batu *et al.* 2013; Levi *et al.* 2013], testing independence of marginals of product distributions [Batu *et al.* 2001] , estimating entropy [Batu *et al.* 2005] are few of the numerous problems that have been studied in the literature. See [Canonne 2015b] for a survey on results related to distribution testing. q Unfortunately, from the modern data analytics point of view, the traditional framework of sampling yields impractical sample complexity. For example, testing if a distribution over a set of n elements is uniform requires $\Omega(\sqrt{n})$ samples from the distribution. The other problems mentioned above have sample complexity at least this high, and in some cases almost linear in n [Raskhodnikova *et al.* 2009; Valiant & Valiant 2011; Valiant 2011].

Conditional Sampling

To remedy this situation, Chakraborty et al. [Chakraborty *et al.* 2016] and Canonne, Ron and Servedio [Canonne *et al.* 2015] proposed a different model, called conditional sampling, which has emerged as a powerful tool for testing properties of probability distributions. In this model, the testers are allowed to sample according to the distribution conditioned on any specific subset of the domain. If the distribution, μ, is over the domain Σ, the tester can submit any subset $S \subseteq \Sigma$ and receive a sample $i \in S$ with probability $\mu(i)/\sum_{j\in S}\mu(j)$, where $\mu(i)$ is the probability of i occurring when a sample is drawn from the distribution μ.

[Canonne *et al.* 2015; Chakraborty *et al.* 2016] proved that in the conditional sampling model, testing uniformity, testing identity to a known distribution, and testing any label-invariant property of distributions are easier than with the unconditional sampling model. Specifically, one can get an algorithm for testing uniformity using $\tilde{O}(1/\epsilon^2)$ conditional samples (conditioning on arbitrary subsets of size 2) [Canonne *et al.* 2015] . Falahatgar et al. [Falahatgar *et al.* 2015], improving an upper bound of $\tilde{O}(1/\epsilon^4)$ in [Canonne *et al.* 2015], showed that testing identity to a known distribution could also be done using $\tilde{O}(1/\epsilon^2)$ conditional samples. They also showed that there exists an algorithm to test identity between two unknown distributions on Σ using $\tilde{O}(\log\log|\Sigma|/\epsilon^5)$ conditional samples. In [Acharya *et al.* 2015a], Acharya, Canonne, and Kamath showed a lower bound of $\Omega(\sqrt{\log\log n})$ for testing equivalence of two unknown distributions.

The sample complexity, in the conditional sampling model, depends on the structure of the condition, *i.e.*, the structure of the subsets (of the domain) on which the distribution is conditioned for drawing samples. Naturally, if there is no restriction on the condition, the tester can sample conditioned on arbitrary subsets, the sample complexity improves. In [Canonne *et al.* 2015], the authors presented an algorithm for testing whether a distribution over $\{1,\ldots,n\}$ is uniform, with sample complexity $\tilde{\Theta}(1/\epsilon^2)$ when conditioning on arbitrary subsets of size 2. However, when the condition set was structured and restricted to intervals, they proved a lower bound of sample complexity $\Omega\left(\frac{\log n}{\log\log n}\right)$. In [Canonne 2015a], Canonne showed that conditioning on interval improves the query complexity of monotonicity testing. Hence it is important to consider the plausible restrictions on the conditions, arising from the structure of the domain.

While [Canonne *et al.* 2015] studied some of the restrictions of the conditions there are many more restrictions, which arise from the structure of the domain and/or arise from other applications, which are yet to be studied. One such important case is when the domain is a Cartesian product of set and one is allowed to condition on the Cartesian product of subsets, but not on arbitrary subsets of the domain.

Testing Joint Distributions: Subcube Conditioning. In practice, data are often multidimensional. In Cryptography, the keys are often defined over $\{0,1\}^n$. Solutions to SAT formulae are over $\{0,1\}^n$ as well. On the other hand, the Lottery Tickets are defined over $[m]^n$ for some $m \in \mathbb{N}$ (each ticket contains n numbers, each from the set $[m]$). In fact, data analysts often get data of million dimensions (features). With higher dimension, comes the "curse of dimensionality." The sample complexity of the testers is exponential in dimension [Acharya *et al.* 2015b; Batu *et al.* 2001; Diakonikolas & Kane 2016], prohibiting practical applications. Very recently, [Daskalakis *et al.* 2018] considered testing higher dimensional *structured* distributions modeled using Markov Random Fields and achieved polynomial (in the dimension) sample complexity under the Ising model. [Canonne *et al.* 2017; Daskalakis & Pan 2017] considered testing properties of *structured* distributions using the probabilistic graphical model and achieved sublinear complexity for certain properties of Bayesian networks. However, all these results assume the distribution is structured and has certain properties. But for arbitrary distributions, testing with practical complexity remains a big concern.

One can indeed be hopeful that using conditional sampling, testing properties of *arbitrary* joint distributions with practical complexity can be achieved. In that case, the assumptions are imposed on the sampling model. Finding correct and natural sampling model is a challenge in itself. While joint distributions can also be viewed as a distribution over a larger domain, the domains of the marginals may be different. Hence sampling conditioned on arbitrary subsets (as used in [Canonne *et al.* 2015; Chakraborty *et al.* 2016]) may not be feasible in real life.

In [Canonne *et al.* 2015], authors also considered structured conditioning, namely Icond (conditioning over an interval) and PCond (conditioning over a pair of points). Icond requires the domain to be well ordered. Moreover, for both the cases, one should be able to sample from arbitrary intervals. For a joint distribution, the natural ordering of the domain is a pair; it involves ordering in the dimensions coupled with ordering in the *individual* domains. For such an ordering, an arbitrary interval required for the Icond tester need not be succinctly encodable and remains impractical.

1.1 Our Results

In this paper, we propose the subcube conditioning model and analyze property testing of joint distributions in that model.

Informally, the subcube conditioning model can be described in the following way. Let Σ^n be the domain of the distribution μ. The Subcube Conditioning Oracle accepts $A_1, A_2, \cdots, A_n \subseteq \Sigma$ and constructs $S = A_1 \times A_2 \times \cdots \times A_n$ as the condition set. The oracle returns a vector $x = (x_1, x_2, \cdots, x_n)$, where each $x_i \in A_i$, with probability $\mu(x)/(\sum_{w \in S} \mu(w))$. If $\mu(S) = 0$, we assume the oracle returns an element from S uniformly at random. We will call these kinds of sample subcube-conditional-samples and call the corresponding sample complexity subcube-conditional-sample complexity. There is no restriction on the individual A_is. *They may be unstructured or structured as pairs or intervals as used in [Canonne* et al. *2015; Chakraborty* et al. *2016].*

Motivation of SubCube conditioning. We believe the subcube conditional sampling model is mathematically interesting in itself. Every Boolean function can be modeled as a subgraph of a hypercube. Testing a property of a Boolean function translates to testing some property of the resulting subgraph. The conditional sampling model is equivalent to sampling over the edges of such subgraph, i.e., to fixing some vertices, sampling over the edges, and checking the properties of the adjacent vertices. We argue sampling over the hypercube arises naturally in many areas.

Database Query. A typical "SELECT" query to a database often looks like `SELECT field1 WHERE field2= cond1 and field2 = cond2`. The response to such a query is all the tuples which satify `cond1` and `cond2`. Sampling over such tuples are indeed conditional sampling.

Side Channel Cryptanalysis. In modern Cryptography, schemes are often "proven" secure (no efficient attack algorithm exists) under the assumption that the keys, internal randomness, and internal memory are inaccessible to the adversary. However, in practice, Cryptographic schemes are deployed in a wide variety of devices, specifically hand-held devices, and smart cards. This situation leads to the "side channel attacks" where tampering with the keys, or internal randomness is feasible. Specifically, the cryptanalytic techniques of fault attacks fix/modify some bits and test the resulting distributions. The subcube conditioning model captures this attack scenario (fixing some bits and testing on resulting subcube).

Our results in this paper can be viewed as proof that "indistinguishability" with uniform (in fact any known distribution) cannot be proven if adversary can tamper with the internal state. [1]

Verification of Random SAT solutions. In software verification and related areas, random solutions to SAT problems are often used as a backbone. However testing whether the solution that one algorithm generates is indeed uniform is a very important problem. Unfortunately, the standard algorithms require impractical complexity. Recently, Chakraborty et al. [Chakraborty & Meel 2016] used the conditional sampling model to get a practically deployable solution. The model of subcube conditioning would be very effective to this problem as one natural conditioning technique is to fix some variables of the SAT equation and then test the distribution of the provided solution.

Recently [Gouleakis *et al.* 2017], has obtained significant improvement in the runtimes of sublinear algorithms for k means clustering and weight estimation of minimum spanning tree using conditional samples. We believe the subcube conditioning can be used in this setting as well.

We remark that the idea of subcube conditioning has been mentioned in the literature related to property testing as well. In fact, analysis of joint distributions using subcube conditioning was posed as a natural open problem in [Canonne *et al.* 2015].

Our Results. We focus on four fundamental properties of distributions: given two joint distributions μ and μ' over Σ^n we would like to test, using subcube-conditional-samples, if (a) μ is uniform, (b) μ is identical to μ' (when μ' is known in advance), (c) μ is identical to μ' (when μ' is not known in advance and has to be accessed using conditional samples), and (d) μ is a product distribution. We have the following four theorems:

Theorem 1.1. *(Informal) Let μ be a probability distribution over Σ^n. There exists an algorithm for testing if μ is uniform, using $\tilde{O}(n^2/\epsilon^2)$ subcube-conditional-samples.*[2]

Theorem 1.2. *(Informal) Let μ be a known probability distribution over the set Σ^n. Let μ' be an unknown distribution over Σ^n. There exists an algorithm to test identity of μ' with μ using $\tilde{O}(n^2/\epsilon^2)$ subcube-conditional-samples.* [2]

Theorem 1.3. *(Informal) Let μ, μ' be unknown distributions over Σ^n. There exists an algorithm to test if μ' and μ are identical using $\tilde{O}(n^5 \log\log|\Sigma|/\epsilon^5)$ subcube-conditional-samples from both μ and μ'.* [2]

Theorem 1.4. *(Informal) Let μ be a probability distribution over the set Σ^n. There exists an algorithm to test whether μ is a product distribution using $\tilde{O}(n^5 \log\log|\Sigma|/\epsilon^5)$ subcube-conditional-samples.* [2]

Comparison to Previous Results. While conditional sampling has been studied in a number of articles in the recent past and although subcube conditioning is a very natural model (that is also

[1] While this result is folklore in Cryptography, the subcube conditioning may be considered as the benchmark model while analyzing the efficiency of a fault attack.

[2] $\tilde{O}$ hides a polynomial function of $\log n$ and $\log(1/\epsilon)$.

discussed in [Canonne *et al.* 2015]), as far as we understand, this is the first formal study on subcube conditioning. One of the main reason for the lack of literature in this area is that the classical setting was not well studied either, till recently. Recently in [Canonne *et al.* 2017] Canonne et al. studied the problem of testing properties of joint distributions over the domain Σ^n. For example, for the fundamental problem of testing if the distribution is uniform, they observed that if the distribution is a product distribution (that is the n marginals are independent), then one needs $\Theta(\sqrt{n})$ samples. But if the distibutions are not independent then in the worst case $\Theta(\Sigma^{n/2})$ samples are necessary.

In comparison, we show that only $\tilde{O}(n^2)$ subcube-conditional samples are necessary in the worst case, and so we have an exponential improvement in the sample complexity. Also, it is interesting to note that the sample complexity for uniformity testing in the subcube model is independent of $|\Sigma|$. This shows that power of subcube conditional samples and also gets the query complexity to a more practical level. Also, from [Canonne *et al.* 2017] we know that $\Omega(\sqrt{n})$ conditional samples are necessary since in the case of product distributions conditional samples give no additional power over standard samples.

A list of our results and comparison to previous results on standard sampling algorithms are given in Table 1.

Problems	Conditional Sampling		Traditional Sampling
	Upper Bound [This paper]	Lower Bound	Upper and Lower Bound
Identity to the uniform distribution	$\tilde{O}(n^2/\epsilon^2)$	$\Omega(\sqrt{n}/\epsilon^2)$ [Canonne *et al.* 2017]	$\Theta(\lvert\Sigma\rvert^{n/2}/\epsilon^2)$ [Paninski 2008]
Identity to a known distribution	$\tilde{O}(n^2/\epsilon^2)$	$\Omega(\sqrt{n}/\epsilon^2)$ [Canonne *et al.* 2017]	$\Theta(\lvert\Sigma\rvert^{n/2}/\epsilon^2)$ [Valiant & Valiant 2014]
Identity between two unknown distributions	$\tilde{O}(n^5 \log\log \lvert\Sigma\rvert/\epsilon^5)$	$\Omega\left(max\left(\sqrt{n}/\epsilon^2, n^{3/4}/\epsilon\right)\right)$ [Canonne *et al.* 2017]	$\Theta\left(\max(\lvert\Sigma\rvert^{2n/3}/\epsilon^{4/3}, \lvert\Sigma\rvert^{n/2}/\epsilon^2)\right)$ [Chan *et al.* 2014]
Identity to a product distribution	$\tilde{O}(n^5 \log\log \lvert\Sigma\rvert/\epsilon^5)$	$\Omega\left(max\left(\sqrt{n}/\epsilon^2, n^{3/4}/\epsilon\right)\right)$ [Canonne *et al.* 2017]	$\Theta(\lvert\Sigma\rvert^{n/2}/\epsilon^2)$ [Acharya *et al.* 2015a] [Diakonikolas & Kane 2016]

Table 1. Comparison between sample complexity of testing joint distributions in Traditional Sampling Model and Subcube Conditioning Model.

Overview of Our Technique. Let us start with the problem of testing if a given distribution is uniform. Let μ be a distribution over Σ^n with marginals $\mu_1, \ldots, \mu_n$.

The simplest case is when μ is a product of n independent distributions. That is, μ_i's are independent but not necessarily identical. But if μ is ϵ-far from uniform one expects to find at least one μ_i which is ϵ/n-far from uniform. Then one can use any tester over Σ if μ_i is far from uniform, which should make at most poly(n) traditional queries. In fact, when μ is a product distribution over $\{0,1\}^n$, [Canonne *et al.* 2017] show that the uniformity and identity can be tested using $O(\sqrt{n}/\epsilon^2)$ unconditional samples. As the marginals of μ are independent and over $\{0,1\}^n$, subcube-conditional-sampling is equivalent to unconditional sampling followed by projections, and hence subcube-conditional samples do not give any additional power in this setting.

But if the μ_i's are not independent then it is possible that all the individual marginals are uniform, but still, the μ is ϵ-far from uniform. As has been observed in [Canonne *et al.* 2017], any algorithm (using unconditional sampling) require exp(n) queries. To circumvent this barrier, we need to use conditional samples. We define a notion of "conditional distance". We show that there exists at least one $i \in [n]$ such that the expected "conditional distance" of ith marginal from uniform is more than $\epsilon/\text{poly}(n)$. Thus it is enough to test for all i if the ith marginal is $\epsilon/\text{poly}(n)$-far from uniform.

We can use the testers from [Canonne *et al.* 2015; Chakraborty *et al.* 2016] to test exactly that using poly(n) subcube-conditional samples. The central idea of the correctness of the algorithm is the correct definition of the "conditional distance" and the "chain rule" that proves that such an i exists. Although the proof of the "chain rule" (given in Section 3) is simple in hindsight, it is a powerful tool that acts as the central backbone for all our upper-bound proofs. Moreover, it gives the flexibility of using adaptive or non adaptive tester over Σ.

1.2 Organization of the paper

In Section 2, we define the notion of conditional distance and SubCube Conditioning. The chain rule is described in Section 3. In Section 4 we present the identity testers and the derived uniformity tester. In Section 5, the tester for testing identity between two unknown distributions is presented. In Section 6, the tester for the independence of marginals is described. In Appendix A we present a lower bound of $n^{1/4}$ for testing identity to the uniform distribution. This lower bound was proved independently of [Canonne *et al.* 2017] and although our lower bound is weaker than their lower bound of $\sqrt{n}$, we feel that our techniques can be of independent interest.

2 NOTATIONS AND PRELIMINARIES

If S is a set $|S|$ denotes the size of the set. If x is a vector of length n, x_i denotes the i^{th} element of x. $x^{(i)}$ denotes the substring of first i elements of x; $x^{(i)} = (x_1, x_2, \cdots, x_i)$. We denote the n-th harmonic number by $H(n)$.

For any set Ω we denote by $\mathcal{U}_\Omega$ the uniform distribution with support Ω. In most of the cases the support of the distribution would be clear from the context and in that case we would drop the subscript and use $\mathcal{U}$ as the uniform distribution over the support in question.

If μ is a distribution with support Ω, for any $x \in \Omega$ we will denote by $\Pr_\mu(x)$ the probability the x occurs when a random sample is drawn from Ω according to μ. If μ is a joint distribution, μ_i denotes the i^{th} marginal distribution of μ.

If μ is a distribution over Σ^n with the marginals $\mu_1, \ldots, \mu_n$ and if the marginals are independent (that is, μ is a product distribution) then we would write $\mu = \mu_1 \otimes \cdots \otimes \mu_n$.

Total Variation Distance. Let μ, μ' be two distributions with support Ω. The variation distance between μ and μ' denoted by $d(\mu, \mu')$ is defined as

$$d(\mu, \mu') := \frac{1}{2} \sum_{x \in \Omega} \left| \Pr_{\mu}(x) - \Pr_{\mu'}(x) \right|.$$

We say μ and μ' are ϵ-*far* (or μ is ϵ-far from μ'), when $d(\mu, \mu') \geq \epsilon$.

If μ is a distribution with support Ω and $A \subseteq \Omega$, then by $(\mu \mid A)$ we denote the distribution over the support A. For any $x \in A$ the probability that x occurs when a random sample is drawn from A (according to the distribution $(\mu \mid A)$) is given by

$$\Pr_{\mu|A}(x) = \frac{\Pr_\mu(x)}{\sum_{y \in A} \Pr_\mu(y)}.$$

Hellinger Distance. Let μ, μ' be two distributions with support Ω. The Hellinger distance between μ and μ' denoted by $H(\mu, \mu')$ is defined as

$$H(\mu, \mu') = \frac{1}{\sqrt{2}} \sqrt{\sum_{x \in \Omega} \left(\sqrt{\Pr_\mu(x)} - \sqrt{\Pr_{\mu'}(x)} \right)^2} = \sqrt{\left(1 - \sum_{x \in \Omega} \sqrt{\Pr_\mu(x) \Pr_{\mu'}(x)} \right)}$$

Hellinger distance has some nice properties and is useful for bounding lower and upper bounding variation distance.

$$d(\mu,\mu') \le 2H(\mu,\mu') \le 2\sqrt{d(\mu,\mu')}$$

Also for any two product distributions $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ and $\mu' = \mu'_1 \otimes \cdots \otimes \mu'_n$

$$H(\mu,\mu')^2 \le \sum_{i=1}^{n} H(\mu_i,\mu'_i)^2.$$

Conditional Distance. Let μ,μ' be two distributions over Ω. Let $A \subseteq \Omega$. The variation distance between μ and μ' conditioned on A (denote by $d(\mu,\mu'|A)$) is defined as

$$d(\mu,\mu'|A) := \frac{1}{2}\sum_{x\in\Omega}\left|\Pr_{\mu|A}(x) - \Pr_{\mu'|A}(x)\right|.$$

We say μ and μ' are ϵ-*far*, conditioned on A, when $d(\mu,\mu'|A) \ge \epsilon$.

Subcube Conditioning. In this paper we work with joint distributions; $\Omega = \Sigma^n$ for some set Σ. We consider conditional distance under the condition on $A = A_1 \times A_2 \times \cdots \times A_n$ where each $A_i \subseteq \Sigma$.

Let μ be a distribution over Σ^n and $X = (X_1, X_2, \ldots, X_n)$ be a random variable distributed according to μ. $\mu^{(i)}$ denotes the distribution over Σ^i where for every $x \in \Sigma^i$,

$$\Pr_{\mu^{(i)}}(x) = \Pr_{X\sim\mu}[(X_1, X_2, \cdots, X_i) = (x_1, x_2, \cdots, x_i)].$$

Let $w \in \Sigma^j$ for some $j < i$. $\mu_i \mid w$ denotes the marginal distribution μ_i when first j random variables fixed to w.

$$\Pr_{\mu_i|w}(x) = \Pr_{X\sim\mu}[X_i = x | \bigwedge_{k=1}^{j} X_k = w_k].$$

Definition 2.1. Let μ,μ' be two distributions over Σ^n. The *conditional marginal distance* of μ_i and μ_i conditioned on w is given by

$$d(\mu_i,\mu'_i \mid w) = \frac{1}{2}\sum_{x\in\Sigma}\left|\Pr_{\mu_i|w}(x) - \Pr_{\mu'_i|w}(x)\right|$$

The *average conditional distance* between μ_i and μ'_i is defined by

$$\mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i,\mu'_i|w)] = \sum_{w\in\Sigma^{i-1}} \Pr_{\mu^{i-1}}(w) d(\mu_i,\mu'_i|w).$$

The SubCube Condition Model. Let μ be a distribution over Σ^n. A *subcube conditional oracle for* μ, denoted $\textsc{SubCond}_\mu$, takes as input a sequence of sets $\{A_i\}_{i\in[n]}$, $A_i \subseteq \Sigma$. Let A be the product set $A_1 \times \cdots \times A_n$. The oracle returns an element $x \in \Sigma^n$ with probability $\frac{\Pr_\mu[x]}{\sum_{x\in A}\Pr_\mu[x]}$ independently of all previous calls to the oracle.

An (ϵ,δ)-SubCond tester for a property $\mathcal{P}$ with conditional sample complexity t is a randomized algorithm, that receives $0 < \epsilon,\delta < 1$, $n \in \mathbb{N}$ and oracle access to $\textsc{SubCond}_\mu$, and operates as follows.

(1) In every iteration, the algorithm (possibly adaptively) generates a set $A = A_1 \times A_2 \times \cdots \times A_n \subseteq \Sigma^n$, based on the transcript and its internal coin tosses, and calls the conditional oracle with A to receive an element x, drawn according to the distribution μ conditioned on A.

(2) Based on the received elements and its internal coin tosses, the algorithm accepts or rejects the distribution μ.
(3) The algorithm makes at most t queries to $\textsc{SubCond}_\mu$, where t can depend on ϵ, δ, Σ and n.

If μ satisfies $\mathcal{P}$ then the algorithm must accept with probability at least $1-\delta$, and if μ is ϵ-far from all distributions satisfying $\mathcal{P}$, then the algorithm must reject with probability at least $1-\delta$.

We will call such a tester an (ϵ, δ)-SubCond $\mathcal{P}$-tester. For example an (ϵ, δ)-SubCond Uniformity-tester is an (ϵ, δ)-SubCond tester that tests if the given distribution is uniform, an (ϵ, δ)-SubCond Identity-tester is an (ϵ, δ)-SubCond tester that tests if the given distribution is identical to a known distribution and an (ϵ, δ)-SubCond Product-tester is an (ϵ, δ)-SubCond tester that tests if the given distribution is a product distribution or far from all the product distributions.

3 CHAIN RULE OF CONDITIONAL DISTANCES

Let μ and μ' be two distributions over Σ^n, and let $X = (X_1, X_2, \ldots, X_n)$ and $X' = (X'_1, X'_2, \ldots, X'_n)$ be the corresponding random variables. For any $1 \le i \le n$, we denote by μ_i and μ'_i the distributions of the ith marginals of μ and μ' respectively.

Lemma 3.1 (Chain Rule of Conditional Distances). *Let μ and μ' be two distributions over Σ^n, and let $X = (X_1, X_2, \ldots, X_n)$ and $X' = (X'_1, X'_2, \ldots, X'_n)$ be two random variables with distribution μ and μ' respectively. Then the following holds.*

$$d(\mu, \mu') \le d(\mu_1, \mu'_1) + \sum_{i=2}^{n} \mathbb{E}_{w \sim \mu^{(i-1)}}[d(\mu_i, \mu'_i | w)]$$

Proof of Lemma 3.1: Let $w = (w_1, w_2, \ldots, w_n) \in \Sigma^n$.
Let $2 \le i \le n$. Recall that $w^{(i)}$ denotes the substring of first i elements of w.

$$\begin{aligned}
2d(\mu^{(i)}, \mu'^{(i)}) &= \sum_{w \in \Sigma^i} |\Pr_{\mu^{(i)}}(w) - \Pr_{\mu'^{(i)}}(w)| \\
&= \sum_{w \in \Sigma^i} |\Pr_{X \sim \mu}[\wedge_{j=1}^{i-1} X_j = w_j] \Pr_{X \sim \mu}[X_i = w_i | \wedge_{j=1}^{i-1} X_j = w_j] \\
&\qquad - \Pr_{X' \sim \mu'}[\wedge_{j=1}^{i-1} X'_j = w_j] \Pr_{X' \sim \mu'}[X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j]| \\
&\le \sum_{w \in \Sigma^i} \left| \Pr_{X \sim \mu}[\wedge_{j=1}^{i-1} X_j = w_j] \left(\Pr_{X \sim \mu}[X_i = w_i | \wedge_{j=1}^{i-1} X_j = w_j] - \Pr[X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j] \right) \right| \\
&\quad + \sum_{w \in \Sigma^i} \left| \Pr[X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j] \left(\Pr_{X \sim \mu}[\wedge_{j=1}^{i-1} X_j = w_j] - \Pr_{X' \sim \mu'}[\wedge_{j=1}^{i-1} X'_j = w_j] \right) \right|
\end{aligned}$$

Now, the second term reduces to,

$$\begin{aligned}
&\sum_{w\in\Sigma^i} \left|\Pr[X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j]\left(\Pr_{X\sim\mu}[\wedge_{j=1}^{i-1} X_j = w_j] - \Pr_{X'\sim\mu'}[\wedge_{j=1}^{i-1} X'_j = w_j]\right)\right| \\
&= \sum_{w'\in\Sigma^{i-1}} \left|\Pr_{X\sim\mu}[\wedge_{j=1}^{i-1} X_j = w'_j] - \Pr_{X'\sim\mu'}[\wedge_{j=1}^{i-1} X'_j = w'_j]\right| \sum_{w_i\in\Sigma} \Pr_{X'\sim\mu'}[X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w'_j] \\
&= \sum_{w'\in\Sigma^{i-1}} \left|\Pr_{X\sim\mu}[\wedge_{j=1}^{i-1} X_j = w'_j] - \Pr_{X'\sim\mu'}[\wedge_{j=1}^{i-1} X'_j = w'_j]\right| \\
&= \sum_{w'\in\Sigma^{i-1}} |\Pr_{\mu^{(i-1)}}(w') - \Pr_{\mu'^{(i-1)}}(w')| \\
&= 2d(\mu^{(i-1)}, \mu'^{(i-1)}).
\end{aligned}$$

The second equality follows from the fact that for each $w' \in \Sigma^{i-i}$, $\sum_{w_i\in\Sigma} \Pr[X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w'_j] = 1$.[3] Hence,

$$d(\mu^{(i)}, \mu'^{(i)}) \le d(\mu^{(i-1)}, \mu'^{(i-1)}) + \sum_{w\in\Sigma^{i-1}} \Pr_{\mu^{(i-1)}}(w) d(\mu_i, \mu'_i | w)$$

Solving the recursion we get the lemma. □

Arranging the marginals by the increasing order of the average conditional distance, we get the immediate corollary.

Lemma 3.2. *If $d(\mu, \mu') \ge \epsilon$, then there exists a $c \le \lceil \log n \rceil$ such that*

$$2^{c-1} \le \left|\left\{i \in [n] \mid \mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i, \mu'_i | w)] \ge \frac{\epsilon}{2^c H(n)}\right\}\right|$$

Proof of Lemma 3.2. Without loss of generality let $i_1, i_2, \ldots, i_n$ be indices such that

$$\mathbb{E}_{w\sim\mu^{(i_1-1)}}[d(\mu_{i_1}, \mu'_{i_1} | w)] \ge \mathbb{E}_{w\sim\mu^{(i_2-1)}}[d(\mu_{i_2}, \mu'_{i_2} | w)] \ge \mathbb{E}_{w\sim\mu^{(i_n-1)}}[d(\mu_{i_n}, \mu'_{i_n} | w)]$$

We will need the following claim.

Claim 3.3. *There exists $k \in [n]$ such that*

$$\mathbb{E}_{w\sim\mu^{(i_k-1)}}[d(\mu_{i_k}, \mu'_{i_k} | w)] \ge \epsilon/(kH(n))$$

Let k be the index from Claim 3.3. We put $c = \lceil \log k \rceil$ to get $\epsilon/2^c H(n) \le \epsilon/kH(n)$. Clearly

$$\left|\left\{i \in [n] \mid \mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i, \mu'_i | w)] \ge \frac{\epsilon}{2^c H(n)}\right\}\right| \ge k \ge 2^{c-1}.$$

□

Proof of Claim 3.3. If no such k exists, then

$$d(\mu, \mu') \le \sum_{k=1}^{n} \mathbb{E}_{w\sim\mu^{(i_k-1)}}[d(\mu_{i_k}, \mu'_{i_k} | w)] < \sum_{k=1}^{n} \epsilon/(kH(n)) \le \epsilon$$

which contradicts the distance assumption in Lemma 3.2. □

[3] If w' is outside of support of μ', like in [Chakraborty *et al.* 2016], we can define the conditional probability to be uniform over Σ

4 TESTING IDENTITY WITH A KNOWN DISTRIBUTION

In this section, we present an identity tester of Sample complexity $\tilde{O}(n^2/\epsilon^2)$. We recall the following result proved in [Falahatgar *et al.* 2015].

LEMMA 4.1. *[Falahatgar* et al. *2015] Let μ be a known distribution over Σ. Given $0 < \epsilon < 1$ and $0 < \delta < 1$ and a distribution μ' over Σ there is an adaptive (ϵ, δ)-SUBCOND Identity Tester with conditional sample complexity $\tilde{O}(\frac{1}{\epsilon^2}\log(\frac{1}{\delta}))$. In other words, there is a tester that draws $\tilde{O}(\frac{1}{\epsilon^2}\log(\frac{1}{\delta}))$ conditional samples and*

- *if $\mu = \mu'$, then the tester will accept with probability $(1-\delta)$, and*
- *if $d(\mu, \mu') \geq \epsilon$ then the tester will reject with probability $(1-\delta)$.*

Let μ be a known distribution over Σ^n, μ' be an unknown distribution over Σ^n that can be accessed via $\text{SUBCOND}_{\mu'}$ oracle, and ϵ be the target distance. The following algorithm tests identity of μ' with μ. We use the identity tester BasicIDTester over Σ guaranteed by Lemma 4.1 as a subroutine.

Algorithm 1: The Identity Tester for Joint Distributions

1: $\delta = 1/3$.
2: $\delta' = \delta\epsilon/64n(\log n)^2$
3: **for** $j = 1$ to $\log n + 1$ **do**
4: $\quad \epsilon_j = \epsilon/2^j H(n)$
5: $\quad \ell_j = \log\left(\frac{2^{j+1}H(n)}{\epsilon}\right)$
6: $\quad$ Create a set S_j by sampling, with replacement, $(4n/2^j)$ element from $[n]$ uniformly at random.
7: $\quad$ **for all** $i \in S_j$ **do**
8: $\quad\quad$ **for** $k = 0$ to ℓ_j **do**
9: $\quad\quad\quad \epsilon_{(j,k)} = 2^{k-1}\epsilon_j$
10: $\quad\quad\quad \delta_k = \delta'/(k+3)^2$
11: $\quad\quad\quad$ **for** $t = 1$ to $2^{k+2}(k+3)^2$ **do**
12: $\quad\quad\quad\quad$ Sample $w \sim \mu$. Let $w = (w_1, \cdots, w_n)$.
13: $\quad\quad\quad\quad$ Consider the distribution $\mu_i \mid w^{(i-1)}$.
14: $\quad\quad\quad\quad$ If BasicIDTester$(\mu_i|w^{(i-1)}, \mu'_i|w^{(i-1)}, \epsilon_{(j,k)}, \delta_k)$ rejects, Output REJECT
15: $\quad\quad\quad$ **end for**
16: $\quad\quad$ **end for**
17: $\quad$ **end for**
18: **end for**
19: Output ACCEPT

THEOREM 4.2. *Given any $0 < \epsilon < 1$, Algorithm 1 is an $(\epsilon, \frac{1}{3})$ -SUBCOND Identity Tester for joint distributions with conditional sample complexity of $\tilde{O}(n^2/\epsilon^2)$ where $\tilde{O}$ hides a polynomial function of $\log n, \log\frac{1}{\epsilon}$.*

NOTE 4.3. *For any $0 < \epsilon, delta < 1$, one can obtain an (ϵ, δ) -SUBCOND Identity Tester by standard techniques of error reduction. The query complexity would increase by a factor of $\log(1/\delta)$.*

4.1 Proof of Theorem 4.2

Fix $\delta = \frac{1}{3}$. In Algorithm 1, Step 14 queries BasicIDTester. BasicIDTester needs conditional samples for testing whether $d(\mu_i, \mu'_i \mid w^{(i-1)}) \geq \epsilon_{(j,k)}$. To answer a conditional query with condition $B \subseteq \Sigma$ for the distribution $\mu'_i|w^{(i-1)}$, we set $A_j = \{w_j\}$ for $j = 1, 2, \ldots, i-1$, $A_i = B$, and $A_j = \Sigma$ for

$j = i+1, \ldots, n$, and query the SubCond oracle with the condition A. This correctly simulates the conditional oracle required by the underlying identity tester. Thus Algorithm 1 is a SubCond Tester.

4.1.1 Sample Complexity of Algorithm 1. By Lemma 4.1, a query to BasicIDTester with parameters $(\mu_i|w^{(i-1)}, \mu'_i|w^{(i-1)}, \epsilon_{(j,k)}, \delta_k)$, requires $\tilde{O}(1/\epsilon^2_{(j,k)})$ samples. Here $\tilde{O}$ hides polylogarithmic factors of $|\Sigma|, \epsilon_{(j,k)}$ including the factors due to $\log(1/\delta_k)$.

For each index in S_j, the sample complexity is

$$\sum_{k=0}^{\ell_j} \tilde{O}\left(\frac{1}{\epsilon^2_{(j,k)}}\right) O\left(2^k k^2\right) = \sum_{k=0}^{\ell_j} \tilde{O}\left(\frac{2^k k^2}{2^{2k}\epsilon_j^2}\right) \qquad [\because \epsilon_{(j,k)} = 2^{k-1}\epsilon_j \text{ by step 9}]$$
$$= \sum_{k=0}^{\ell_j} \tilde{O}\left(\frac{k^2}{2^k \epsilon_j^2}\right)$$

Here $\tilde{O}$ hides some polylogarithmic function of k and $1/\epsilon_j$. As $k \le \ell_j = \log\left(\frac{2}{\epsilon_j}\right)$, the expression can be bounded as

$$\sum_{k=0}^{\ell_j} \tilde{O}\left(\frac{k^2}{2^k\epsilon_j^2}\right) = \tilde{O}\left(\frac{1}{\epsilon_j^2}\right) \sum_{k=0}^{\ell_j} O\left(\frac{k^2}{2^k}\right) = \tilde{O}\left(\frac{1}{\epsilon_j^2}\right)$$

The last equality holds true as $\sum_{k\ge 0} \frac{k^2}{2^k} = 6$.

The size of S_j is $\frac{4n}{2^j}$. Adding over all possible j, we get the total sample complexity

$$\sum_{j=1}^{\log n+1} \frac{4n}{2^j}\tilde{O}\left(\frac{1}{\epsilon_j^2}\right) = \sum_{j=1}^{\log n+1} \frac{4n}{2^j}\tilde{O}\left(\frac{2^{2j}H(n)^2}{\epsilon^2}\right) \qquad [\because \epsilon_j = \epsilon/2^j H(n) \text{ by step 4}]$$
$$= \tilde{O}\left(\frac{nH(n)^2}{\epsilon^2}\right) \sum_{j=1}^{\log n+1} 2^j = \tilde{O}(n^2/\epsilon^2)$$

4.1.2 Correctness of the Algorithm 1. Completeness.We will show that if $d(\mu, \mu') = 0$ is the algorithm will reject with probability at most δ.

Algorithm 1, rejects μ' if there exists $i \in [n]$ and a sampled $w = (w_1, \cdots, w_n) \in \Sigma^n$ the underlying Identity Tester rejects in the Step 14.

Suppose μ and μ' are identical. Then for all $w \in \Sigma^{i-1}$, $\mu_i|w$ is identical to $\mu'_i \mid w$. For each query, BasicIDTester will reject in Step 14 with probability at most δ_k. By union bound, the probability that algorithm will reject μ' is at most

$$\sum_{j=1}^{\log n+1}\frac{4n}{2^j}\sum_{k=0}^{\ell_j}(2^{k+2}(k+3)^2\delta_k) = \sum_{j=1}^{\log n+1}\frac{4n}{2^j}\sum_{k=0}^{\ell_j}(2^{k+2}\delta') \qquad [\because \delta_k = \delta'/(k+3)^2 \text{ by step 10}]$$

$$= 16\delta'\sum_{j=1}^{\log n+1}\frac{n}{2^j}\sum_{k=0}^{\ell_j}2^k$$

$$< 16\delta'\sum_{j=1}^{\log n+1}\frac{n}{2^j}2^{\ell_j+1}$$

$$= 64\delta'\sum_{j=1}^{\log n+1}\frac{n}{2^j}\frac{2^jH(n)}{\epsilon}$$

$$< \frac{64\delta' n(\log n)^2}{\epsilon} = \delta \qquad [\because \delta' = \delta\epsilon/64n(\log n)^2 \text{ by step 2}]$$

Soundness. Now we prove the soundness of the Algorithm 1. Let μ be a distribution over Σ^n and $d(\mu,\mu') \geq \epsilon$. We shall show that Algorithm 1, rejects μ' with probability at least $2/3$.
Let

$$\tau_c \stackrel{def}{=} \left\{i \in [n] \mid \mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i,\mu_i'|w)] \geq \frac{\epsilon}{2^cH(n)}\right\}$$

Let $c \leq \lceil \log n \rceil$ be the integer guaranteed by Lemma 3.2, such that $|\tau_c| \geq 2^{c-1}$. Note, $\ell_c = \lceil \log\left(\frac{2^{c+1}H(n)}{\epsilon}\right)\rceil$. For each $i \in \tau_c$, for each $k \in [\ell_c]\cup\{0\}$ define

$$\Gamma_{i,k} \stackrel{def}{=} \left\{w \in \Sigma^{i-1} \mid d(\mu_i,\mu_i'|\wedge_{j=1}^{i-1}X_j = w_j) < \frac{2^{k-1}\epsilon}{2^cH(n)}\right\}$$

We require the following lemma based on Levin's economical work investment strategy [Goldreich 2017].

Lemma 4.4. *Let μ be a distribution over Σ^n, and μ is ϵ-far from uniform. Let $X = (X_1,\cdots,X_n)$ be a random variable with distribution μ. Let $w = (w_1,w_2,\cdots,w_n)$ be a random sample drawn from Σ^n according to the distribution μ. Let $\epsilon_c = \frac{\epsilon}{2^cH(n)}$ and $\ell_c = \lceil\log\left(\frac{2}{\epsilon_c}\right)\rceil$.*

Then for all $i \in \tau_c$, there exists $k \in [\ell_c]\cup\{0\}$,

$$\Pr_{w\sim\mu}\left[d(\mu_i,\mu_i' \mid w^{i-1}) \geq 2^{k-1}\epsilon_c\right] \geq \frac{1}{2^k(k+3)^2} \tag{1}$$

(Proof of Lemma 4.4.) From Lemma 3.2, for all index $i \in \tau_c$

$$\mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i,\mu_i'|w)] = \sum_{w\in\Sigma^{i-1}}\Pr_{\mu^{i-1}}(w)d(\mu_i,\mu_i'|w) \geq \frac{\epsilon}{2^cH(n)}$$

Fix $i \in \tau_c$. Let us define

$$B_k \stackrel{def}{=} \{w \in \Sigma^{i-1} : 2^{k-1}\epsilon_c \leq d(\mu_i,\mu_i' \mid w) < 2^k\epsilon_c\} \qquad k \in [\ell_c]\cup\{0\}$$

$$B_{-1} \stackrel{def}{=} \{w \in \Sigma^{i-1} : d(\mu_i,\mu_i' \mid w) < \epsilon_c/2\}$$

By construction, $B_{\ell_c+1} = \emptyset$. We shall prove that there exists $k \in [\ell_c]\cup\{0\}$ such that $\Pr_{w\sim\mu}[w \in B_k] \geq \frac{1}{2^k(k+3)^2}$. Suppose, towards contradiction, for all $k \in [\ell_c]\cup\{0\}$, $\Pr_{w\sim\mu}[w \in B_k] < \frac{1}{2^k(k+3)^2}$.

Then

$$\begin{aligned}
\mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i,\mu_i'|w)] &= \sum_{w\in\Sigma^{i-1}} \Pr_{\mu^{i-1}}(w)d(\mu_i,\mu_i'|w) \\
&= \sum_{w\in B_{-1}} \Pr_{\mu^{i-1}}(w)d(\mu_i,\mu_i' \mid w) + \sum_{k\in[\ell_c]\cup\{0\}}\sum_{w\in B_k} \Pr_{\mu^{i-1}}(w)d(\mu_i,\mu_i' \mid w) \\
&< \Pr_{w\sim\mu}[w\in B_{-1}]\frac{\epsilon_c}{2} + \sum_{k\in[\ell_c]\cup\{0\}} \Pr_{w\sim\mu}[w\in B_k]2^k\epsilon_c \\
&< \frac{\epsilon_c}{2} + \sum_{k\in[\ell_c]\cup\{0\}} \frac{1}{2^k(k+3)^2}2^k\epsilon_c \\
&= \frac{\epsilon_c}{2} + \sum_{k\in[\ell_c]} \frac{\epsilon_c}{(k+2)^2} \\
&< \epsilon_c
\end{aligned}$$

In the last inequality we used the fact that $\sum_{k\in[\ell_c]} \frac{1}{(k+2)^2} < \sum_{k\geq 0} \frac{1}{(k+2)^2}$ which is less than $1/2$. □

By Lemma 4.4, there exists $0 \leq k \leq \ell_c$, such that,

$$Pr_{w\sim\mu^{i-1}}[w\in\Gamma_{i,k}] < \left(1 - \frac{1}{2^k(k+3)^2}\right) \tag{2}$$

Let S_j be the set of indices sampled in the Step 3 in the j^{th} iteration. If Algorithm 1 fails to reject μ', one of the following three cases happen.

(1) No index from τ_c was sampled in S_j. Specifically, $S_c \cap \tau_c = \emptyset$. The probability of this event is

$$\left(1 - \frac{|\tau_c|}{n}\right)^{4n/2^c} \leq e^{-2}.$$

(2) For all index $i \in S_c \cap \tau_c$, for each $k \in [\ell_c] \cup \{0\}$, all the sampled w's are from the set $\Gamma_{i,k}$. The probability of this event is

$$\left(1 - \frac{1}{2^k(k+3)^2}\right)^{2^{k+2}(k+3)^2} \leq e^{-4}.$$

(3) For all index $i \in S_c \cap \tau_c$, for each $k \in [\ell_c] \cup \{0\}$, for all the sampled $w \notin \Gamma_{i,k}$, underlying identity tester fails to reject. Probability of such an event is at most δ', which is less than $1/100$ for $n \geq 2$.

Hence, the probability that Algorithm 1 fails to reject μ' is at most $e^{-2} + e^{-4} + 1/100 < 1/3$. This completes the proof of Theorem 4.2. □

4.2 Uniformity Tester for Arbitrary Joint Distribution

If we set μ to be the uniform distribution, then Algorithm 1 gives us a Uniformity Tester. Hence, we get the following as a corollary of Theorem 4.2.

Theorem 4.5. *Given any $0 < \epsilon < 1$, there exists an $(\epsilon, \frac{1}{3})$-SubCond Uniformity Tester for any joint distribution with conditional sample complexity of $\tilde{O}(n^2/\epsilon^2)$ where $\tilde{O}$ hides a polynomial function of $\log n, \log \frac{1}{\epsilon}$.*

5 IDENTITY TESTING BETWEEN UNKNOWN JOINT DISTRIBUTIONS

In this section, we present Algorithm 2, to test identity when both μ and μ' are unknown. The first change, from Algorithm 1, we need to make is in Step 12. In this case we can no longer sample on our own. However, we can query μ to get w. Secondly, instead of Algorithm BasicIDTester, we need to use Algorithm BasicUnknown guaranteed by the following lemma.

LEMMA 5.1. *[Falahatgar* et al. *2015] Given $0 < \epsilon < 1$ and $0 < \delta < 1$ and distributions μ, μ' over Σ there is an (ϵ, δ)-Identity Tester with conditional sample complexity $\tilde{O}(\frac{\log\log|\Sigma|}{\epsilon^5}\log(\frac{1}{\delta}))$. In other words, there is a tester that draws $\tilde{O}(\frac{\log\log|\Sigma|}{\epsilon^5}\log(\frac{1}{\delta}))$ independent conditional samples and*

- *if $\mu = \mu'$, then the tester will accept with probability $(1-\delta)$, and*
- *if $d(\mu, \mu') \geq \epsilon$ then the tester will reject with probability $(1-\delta)$.*

Algorithm 2: The Identity Tester for two Unknown Joint Distributions

1: $\delta = 1/3$.
2: $\delta' = \delta\epsilon/64n(\log n)^2$
3: **for** $j = 1$ to $\log n + 1$ **do**
4: $\quad\epsilon_j = \epsilon/2^j H(n)$
5: $\quad\ell_j = \log\left(\frac{2}{\epsilon_j}\right)$
6: $\quad$Create a set S_j by sampling, with replacement, $(4n/2^j)$ element from $[n]$ uniformly at random.
7: $\quad$**for all** $i \in S_j$ **do**
8: $\quad\quad$**for** $k = 0$ to ℓ_j **do**
9: $\quad\quad\quad\epsilon_{(j,k)} = 2^{k-1}\epsilon_j$
10: $\quad\quad\quad\delta_k = \delta'/(k+3)^2$
11: $\quad\quad\quad$**for** $t = 1$ to $2^{k+2}(k+3)^2$ **do**
12: $\quad\quad\quad\quad$Query oracle μ to get $w \sim \mu$. Let $w = (w_1, \cdots, w_n)$.
13: $\quad\quad\quad\quad$Consider the distribution $\mu_i \mid w^{(i-1)}$.
14: $\quad\quad\quad\quad$If BasicUnknown$(\mu_i|w^{(i-1)}, \mu_i'|w^{(i-1)}, \epsilon_{(j,k)}, \delta_k)$ rejects, Output REJECT
15: $\quad\quad\quad$**end for**
16: $\quad\quad$**end for**
17: $\quad$**end for**
18: **end for**
19: Output ACCEPT

To prove correctness of Algorithm 2, we note that, in the chain rule, the expectation is over only one distribution. Hence it is sufficient to (unconditionally) query only μ to get w, and apply Lemma 3.2. The rest of the proof is exactly the same as in Section 4.

Sample Complexity of Algorithm 2. By Lemma 5.1, each invocation of BasicUnknown with parameter ϵ_k,δ_k requires $\tilde{O}(\log\log|\Sigma|/\epsilon_k^5)$ samples. As in the case for Algorithm 1, for each index in S_j, the sample complexity is $\tilde{O}(\log\log|\Sigma|/\epsilon^5)$. Hence, the total sample complexity of Algorithm

2 is

$$\begin{aligned}\sum_{j=1}^{\log n+1} \frac{4n}{2^j} \times \tilde{O}\left(\frac{\log\log|\Sigma|}{\epsilon_j^5}\right) &= \sum_{j=1}^{\log n+1} \frac{4n}{2^j} \times \tilde{O}\left(\frac{2^{5j}H(n)^5\log\log|\Sigma|}{\epsilon^5}\right)\\ &= \tilde{O}\left(\frac{nH(n)^5\log\log|\Sigma|}{\epsilon^5}\right)\sum_{j=1}^{\log n+1} 2^{4j}\\ &= \tilde{O}\left(\frac{n^5\log\log|\Sigma|}{\epsilon^5}\right)\end{aligned}$$

Theorem 5.2. *Given $0 < \epsilon < 1$, Algorithm 2 is an $(\epsilon, \frac{1}{3})$-SubCond Identity Tester for two unknown joint distributions with conditional sample complexity of $\tilde{O}\left(\frac{n^5\log\log|\Sigma|}{\epsilon^5}\right)$ where $\tilde{O}$ hides a polynomial function of $\log n, \log\frac{1}{\epsilon}$.*

6 TESTING INDEPENDENCE OF MARGINALS

Let μ be a probability distribution over Σ^n. In this section we present an algorithm to test whether μ is a product distribution; i.e., *whether all the marginals of μ are independent or μ is far from all the product distributions.*

Define μ' to be the product of marginals of μ.

$$\Pr_{\mu'}(w) = \prod_{i=1}^{n}\Pr_{\mu_i}(w_i)\ \forall\ w \in \Sigma^n$$

By definition, the marginal distributions μ_i' are exactly the marginal distributions μ_i. If μ is ϵ-far from all the product distributions, it is ϵ-far from μ'. Using the chain rule (Lemma 3.1),

$$\begin{aligned}d(\mu,\mu') &\leq d(\mu_1,\mu_1') + \sum_{i=2}^{n}\mathbb{E}_{w\sim\mu^{(i-1)}}[d(\mu_i,\mu_i'|w)]\\ &= \sum_{i=2}^{n}\sum_{w\in\Sigma^{i-1}} Pr_{\mu^{(i-1)}}(w)\left(\sum_{w_i\in\Sigma}\left|\Pr_{\mu_i}(w_i|w) - \Pr_{\mu_i'}(w_i|w)\right|\right)\\ &= \sum_{i=2}^{n}\sum_{w\in\Sigma^{i-1}} Pr_{\mu^{(i-1)}}(w)d(\mu_i|w,\mu_i)\end{aligned}$$

Therefore, we need to test whether there exists $i \in [n]$, such that the marginal distribution μ_i is far (on average) from the conditional marginal distribution $\mu_i|w$. As both μ_i and $\mu_i|w$ is distributed over Σ, we can again use BasicUnknown tester from [Falahatgar *et al.* 2015], where identity between two unknown distributions is tested using $\tilde{O}(\log\log|\Sigma|/\epsilon^5)$ sample complexity. The only thing left is to sample w according to μ^{i-1}. Such a w can be sampled by taking an unconditionally sampled string and selecting the first $i-1$ bit of that string. Rest of the algorithm is exactly same as in Algorithm 2.

Theorem 6.1. *For any $0 < \epsilon < 1$, there exists an $(\epsilon, \frac{1}{3})$- SubCond Product Tester for joint distributions with conditional sample complexity of $\tilde{O}\left(\frac{n^5\log\log|\Sigma|}{\epsilon^5}\right)$, where $\tilde{O}$ hides a polynomial function of $\log n, \log\left(\frac{1}{\epsilon}\right)$*

The proof of Theorem 6.1 follows directly from Theorem 5.2, and the observation that in this particular case the (conditional) samples for μ_i can be produced by conditioning only on the i^{th} index of Σ^n.

7 CONCLUSION

In this paper we analyzed property testing of joint distributions in the conditional sampling model. We considered the natural subcube conditioning and presented testers to test uniformity, identity with a known distribution, identity with an unknown distribution, and independence of marginals of query complexity polynomial in the dimension, thus avoiding the curse of dimensionality.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for insightful suggestions and comments which significantly improved the paper. In particular the authors would like to thank the first reviewer for suggesting the use of Levin's economic work strategy which resulted in a speedup of all our algorithms by a factor of n/ϵ.

Rishiraj is supported by SERB ECR/2017/001974.

REFERENCES

Acharya, Jayadev, Canonne, Clément L., & Kamath, Gautam. 2015a. A Chasm Between Identity and Equivalence Testing with Conditional Queries. *Pages 449–466 of: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA.*

Acharya, Jayadev, Daskalakis, Constantinos, & Kamath, Gautam. 2015b. Optimal Testing for Properties of Distributions. *Pages 3591–3599 of: Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada.*

Batu, Tuğkan, Fortnow, Lance, Fischer, Eldar, Kumar, Ravi, Rubinfeld, Ronitt, & White, Patrick. 2001 (Oct. 14–17). Testing Random Variables for Independence and Identity. *Pages 442–451 of:* Werner, Bob (ed), *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science (FOCS-01).*

Batu, Tuğkan, Dasgupta, Sanjoy, Kumar, Ravi, & Rubinfeld, Ronitt. 2005. The Complexity of Approximating the Entropy. *SIAM J. Comput.*, **35**(1), 132–150.

Batu, Tuğkan, Fortnow, Lance, Rubinfeld, Ronitt, Smith, Warren D., & White, Patrick. 2013. Testing Closeness of Discrete Distributions. *Journal of the ACM*, **60**(1), 4:1–4:25.

Canonne, Clément L. 2015a. Big Data on the Rise? - Testing Monotonicity of Distributions. *Pages 294–305 of: Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I.*

Canonne, Clément L. 2015b. A Survey on Distribution Testing: Your Data is Big. But is it Blue? *Electronic Colloquium on Computational Complexity (ECCC)*, **22**, 63.

Canonne, Clément L., Ron, Dana, & Servedio, Rocco A. 2015. Testing Probability Distributions using Conditional Samples. *SIAM Journal on Computing*, **44**(3), 540–616.

Canonne, Clément L., Diakonikolas, Ilias, Kane, Daniel M., & Stewart, Alistair. 2017. Testing Bayesian Networks. *Pages 370–448 of: Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017.*

Chakraborty, Sourav, & Meel, Kuldeep. 2016. Testing correctness of programs that claim to produce satisfying assignments uniformly at random. *Under Preparation.*

Chakraborty, Sourav, Fischer, Eldar, Goldhirsh, Yonatan, & Matsliah, Arie. 2016. On the Power of Conditional Samples in Distribution Testing. *SIAM J. Comput.*, **45**(4), 1261–1296.

Chan, Siuon, Diakonikolas, Ilias, Valiant, Paul, & Valiant, Gregory. 2014. Optimal Algorithms for Testing Closeness of Discrete Distributions. *In:* Chekuri, Chandra (ed), *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014.* SIAM.

Daskalakis, Constantinos, & Pan, Qinxuan. 2017. Square Hellinger Subadditivity for Bayesian Networks and its Applications to Identity Testing. *Pages 697–703 of: Proceedings of the 30th Conference on Learning Theory, COLT 2017, Amsterdam, The Netherlands, 7-10 July 2017.*

Daskalakis, Constantinos, Dikkala, Nishanth, & Kamath, Gautam. 2018. Testing Ising Models. *Pages 1989–2007 of: Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018.*

Diakonikolas, Ilias, & Kane, Daniel M. 2016. A New Approach for Testing Properties of Discrete Distributions. *Pages 685–694 of: IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA.*

Falahatgar, Moein, Jafarpour, Ashkan, Orlitsky, Alon, Pichapati, Venkatadheeraj, & Suresh, Ananda Theertha. 2015. Faster Algorithms for Testing under Conditional Sampling. *Pages 607–636 of: Proceedings of The 28th Conference on Learning*

Theory, COLT 2015, Paris, France, July 3-6, 2015.

Fischer, Eldar. 2004. The difficulty of testing for isomorphism against a graph that is given in advance. *Pages 391–397 of: Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004.*

Fischer, Eldar, Newman, Ilan, & Sgall, Jirí. 2004. Functions that have read-twice constant width branching programs are not necessarily testable. *Random Struct. Algorithms*, **24**(2), 175–193.

Goldreich, Oded. 2017. *Introduction to Property Testing.* Cambridge University Press.

Goldreich, Oded, & Ron, Dana. 2011. On Testing Expansion in Bounded-Degree Graphs. *Pages 68–75 of:* Goldreich, Oded (ed), *Studies in Complexity and Cryptography.* Lecture Notes in Computer Science, vol. 6650. Springer.

Goldreich, Oded, Goldwasser, Shafi, & Ron, Dana. 1998. Property Testing and its Connection to Learning and Approximation. *J. ACM*, **45**(4), 653–750.

Gouleakis, Themistoklis, Tzamos, Christos, & Zampetakis, Manolis. 2017. Faster Sublinear Algorithms Using Conditional Sampling. *In: Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms.* SODA '17.

Levi, Reut, Ron, Dana, & Rubinfeld, Ronitt. 2013. Testing Properties of Collections of Distributions. *Theory of Computing*, **9**, 295–347.

Paninski, L. 2008. A Coincidence-Based Test for Uniformity Given Very Sparsely Sampled Discrete Data. *IEEE Trans. Inf. Theor.*, **54**(10), 4750–4755.

Raskhodnikova, Sofya, Ron, Dana, Shpilka, Amir, & Smith, Adam. 2009. Strong Lower Bounds for Approximating Distribution Support Size and the Distinct Elements Problem. *SIAM J. Comput.*, **39**(3), 813–842.

Rubinfeld, Ronitt, & Sudan, Madhu. 1996. Robust Characterizations of Polynomials with Applications to Program Testing. *SIAM J. Comput.*, **25**(2), 252–271.

Valiant, Gregory, & Valiant, Paul. 2011. Estimating the unseen: an $n/\log(n)$-sample estimator for entropy and support size, shown optimal via new CLTs. *Pages 685–694 of: Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011.*

Valiant, Gregory, & Valiant, Paul. 2014. An Automatic Inequality Prover and Instance Optimal Identity Testing. *Pages 51–60 of: 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014.*

Valiant, Paul. 2011. Testing Symmetric Properties of Distributions. *SIAM J. Comput.*, **40**(6), 1927–1968.

A A WEAKER LOWER BOUND WITH SIMPLE PROOF

THEOREM A.1. *For any $0 \leq \epsilon \leq 1/2$ any $(\epsilon, 1/3)$ − SUBCOND Uniformity-Tester has subcube-conditional sample complexity $\Omega(\sqrt[4]{n}/\sqrt{\epsilon})$. The lower bound holds even for the case when the domain is $\{0,1\}^n$ and the given distribution is a product of n independent (though not necessarily identical) distributions.*

PROOF. Let μ be a product distributions over the domain $\{0,1\}^n$ with marginals $\mu_1, \ldots, \mu_n$. So $\mu = \mu_1 \otimes \cdots \otimes \mu_n$. Note that since the μ_i are independent, if $i \neq j$ then conditioning on μ_i does not affect the samples we get from a μ_j. Also since the μ_i are all distributions over a two element set (namely $\{0,1\}$) conditioning on any subset of $\{0,1\}$ also of no use. Thus drawing subcube-conditional-samples from μ is as good as drawn samples (without any conditioning) from μ.

So it is sufficient for us to prove that for any $0 \leq \epsilon \leq 1/2$ any $(\epsilon, 1/3)$ Uniformity-Tester has sample complexity $\Omega(\sqrt[4]{n})$, when the domain is $\{0,1\}^n$ and the given distributions are product distributions.

The main idea of the proof is to use a standard technique from property testing where the following lemma is used. The following lemma has been rewritten in the language and context of this paper. A proof of the general statement of the lemma can be found in [Fischer 2004; Fischer *et al.* 2004].

THEOREM A.2. *Let P be a property of distributions over σ^n that we want to test. Suppose $\mathcal{D}_Y$ is a distribution over all the distributions that satisfy the given property P, and let $\mathcal{D}_N$ be a distribution over all distributions that are ϵ-far from satisfying the property P. Let Q_Y be the distribution over outcomes of q samples when the samples are drawn from a distribution D_Y, that is drawn according according to $\mathcal{D}_Y$. Similarly, let Q_N be the distribution over outcomes of q samples when the samples are*

drawn from a distribution D_N, that is drawn according to the $\mathcal{D}_N$. If the variation distance between Q_Y and Q_N is less than $1/3$ then any $(\epsilon, 1/3)$-Tester for the property P will have sample complexity more than q.

In the context of our theorem we have the property P is "Uniformity". So the distribution $\mathcal{D}_Y$ is the uniform distribution over the domain $\{0,1\}$. Now let us define the distribution $\mathcal{D}_N$:

Let D_1 be the distribution over $\{0,1\}$ where 1 is produced with probability $(1/2 + 2\sqrt{\frac{\epsilon}{n}})$ and 0 produced with probability $(1/2 - 2\sqrt{\frac{\epsilon}{n}})$. And let D_0 be the distribution over $\{0,1\}$ where 1 is produced with probability $(1/2 - 2\sqrt{\frac{\epsilon}{n}})$ and 0 produced with probability $(1/2 + 2\sqrt{\frac{\epsilon}{n}})$.

Consider the set of distributions $\mathcal{D}$ over $\{0,1\}^n$ which are a product of n distribution each of which is either D_0 or D_1. That is,

$$\mathcal{D} = \{\mu_1 \otimes \cdots \otimes \mu_n \mid \text{for all i, } \mu_i \text{ is either } D_0 \text{ or } D_1\}$$

Claim A.3. *Any $\mu \in \mathcal{D}$ is ϵ-far from uniform. That is, for any $\mu \in \mathcal{D}$ we have*

$$d(\mu, \mathcal{U}) \geq \epsilon$$

From Claim A.3 we see that all the distributions in $\mathcal{D}$ are ϵ-far from uniform. Thus we can take the distribution $\mathcal{D}$ as our distribution $\mathcal{D}_N$. If a distribution is drawn from $\mathcal{D}_N$ or $\mathcal{D}_Y$, q samples from the distribution would give q many $\{0,1\}$-strings of length n. Note that if a distribution is drawn from $\mathcal{D}_Y$ (that is, the distribution is the uniform distribution over $\{0,1\}^n$) then the distribution of the outcomes of q samples is a uniform distribution over $\{0,1\}^{nq}$. So, by theorem A.2, it is enough to show that if μ is drawn from $\mathcal{D}_N$ then the distribution of the outcomes (as a distribution over $\{0,1\}^{nq}$) is $1/3$-close to uniform.

Note that μ is a distribution drawn from $\mathcal{D}_N$ we can think of μ as $\mu_1 \otimes \cdots \otimes \mu_n$ where each μ_i is independently and uniformly choose from the set $\{D_0, D_1\}$. Let μ^q be the distribution over $\{0,1\}^{nq}$ when q samples are drawn from μ. And now the following lemma completes the proof of Theorem A.1.

Lemma A.4. *If $q \leq \frac{\sqrt[4]{n}}{20\sqrt{\epsilon}}$ then*

$$d(\mu^q, \mathcal{U}) \leq \frac{1}{3}.$$

□

A.1 Proof of Claim A.3

Let $\mu = \mu_1 \otimes \cdots \otimes \mu_n$. Without loss of generality we will assume that all the μ_i's are the distribution D_1. That is 1 is produced with probability $(1/2 + 2\sqrt{\frac{\epsilon}{n}})$ and 0 produced with probability $(1/2 - 2\sqrt{\frac{\epsilon}{n}})$. For simplifying notations we will assume 1 is produced with probability $(1/2 + \epsilon')$ and 0 produced with probability $(1/2 - \epsilon')$.

Since we know $d(\mu, \mathcal{U}) \geq H(\mu, \mathcal{U})^2$, it is enough for us to prove $H(\mu, \mathcal{U})^2 \geq \epsilon$. For any $x \in \{0,1\}^n$ let $p(x)$ be the probability of getting x when drawn from μ. Note that the probability of getting x when drawn from $\mathcal{U}$ is $1/2^n$.

By definition we have

$$H(\mu, \mathcal{U})^2 = \frac{1}{2} \sum_{x \in \{0,1\}^n} \left(\sqrt{p(x)} - \sqrt{1/2^n}\right)^2 = 1 - \sum_{x \in \{0,1\}^n} \left(\sqrt{p(x)/2^n}\right)$$

Now note that if x has k 1's and $(n-k)$ 0's then $p(x) = (1/2 + \epsilon')^k (1/2 - \epsilon')^{n-k}$. So we have

$$\sum_{x \in \{0,1\}^n} \left(\sqrt{p(x)/2^n}\right) = \frac{1}{2^n} \sum_{k=0}^{n} \binom{n}{k} \sqrt{(1 + 2\epsilon')^k (1 - 2\epsilon')^{n-k}} = \frac{1}{2^n} \left(\sqrt{(1 + 2\epsilon')} + \sqrt{(1 - 2\epsilon')}\right)^n$$

Now since $(\sqrt{1+x}+\sqrt{1-x}) \leq 2(1-\frac{x^2}{8})$ for all $x \leq 1$ so,

$$\frac{1}{2^n}\left(\sqrt{(1+2\epsilon')}+\sqrt{(1-2\epsilon')}\right)^n \leq \left(1-\frac{\epsilon'^2}{2}\right)^n \leq \left(1-\frac{\epsilon'^2 n}{2}+\frac{\epsilon'^4}{4}\binom{n}{2}\right).$$

The last inequality follows from the fact that $(1-x)^n \leq (1-xn+\binom{n}{2}x^2)$. Now putting all the things together we have

$$H(\mu,\mathcal{U})^2 \geq \left(1-\left(1-\frac{\epsilon'^2 n}{2}+\frac{\epsilon'^4}{4}\binom{n}{2}\right)\right) \geq \left(\frac{\epsilon'^2 n}{2}-\frac{\epsilon'^4}{4}\binom{n}{2}\right)$$

If $\epsilon' = 2\sqrt{\epsilon/n}$ then from the above inequality, and the fact that $\epsilon < 1/2$, we have $H(\mu,U)^2 \geq \epsilon$.

A.2 Proof of Lemma A.4

Let us start with a claim. We defer the proof of the claim to the end of this section.

Claim A.5. *If P and Q be two distributions over Σ and for all $x \in \Sigma$ we have*

$$\Pr_P(x) = (1+\epsilon_x)\Pr_Q(x)$$

then we have

$$H(P,Q)^2 \leq \frac{1}{2}\sum_{x\in\Sigma}\epsilon_x^2 \Pr_Q(x)$$

Claim A.5 helps to upper bound the Hellinger distance in terms of the ℓ_∞ distance. Now let $\Sigma = \{0,1\}^q$. And let μ_i^q be the distribution on Σ that is obtained by drawing q samples from μ_i. Clearly, $\mu^q = \mu_1^q \otimes \mu_2^q \otimes \cdots \otimes \mu_n^q$. To prove that the variation distance of μ^q from uniform is less than $1/3$ we will first show that the ℓ_∞ distance of μ_i from uniform is small then using Claim A.5 we get that the Hellinger distance of μ_i^q from uniform is small.. And then we can show that if all the μ_i^q has small Hellinger distance from uniform then μ^q has small Hellinger distance from uniform which would give an upper bound on the variation distance of μ^q from uniform.

Now the following claim upper bounds the ℓ_∞ distance of μ_i^q from uniform.

Claim A.6. *For all i and for all $x \in \Sigma$*

$$|\Pr_{\mathcal{U}}(x) - \Pr_{\mu_i^q}(x)| \leq \frac{10\epsilon q^2}{2^q n}$$

Or in other words for all $x \in \Sigma$ if

$$\Pr_{\mu_i^q}(x) = (1\pm\epsilon_x)\Pr_{\mathcal{U}}(x)$$

then $|\epsilon_x| \leq 10\epsilon q^2/n$

By definition of Hellinger distance and variation distance we have

$$d(\mu^q,\mathcal{U}) = \sum_{x\in\{0,1\}^{qn}}\left|\Pr_{\mu^q}(x)-\Pr_{\mathcal{U}}(x)\right| \leq 2H(\mu^q,\mathcal{U})$$

Again we know that for any two product distributions $P = P_1 \otimes \cdots \otimes P_n$ and $Q = Q_1 \otimes \cdots \otimes Q_n$

$$H(P_1\otimes\cdots\otimes P_n, Q_1\otimes\cdots\otimes Q_n)^2 \leq \sum_{i=1}^n H(P_i,Q_i)^2.$$

Thus we have

$$d(\mu^q,\mathcal{U}) \leq 2\sqrt{\left(\sum_{i=1}^n H(\mu_i^q,\mathcal{U})^2\right)} \tag{3}$$

From Equation 3 and Claim A.5 we have

$$d(\mu^q, \mathcal{U}) \leq 2\sqrt{\sum_{i=1}^{n} \frac{1}{2} \sum_{x \in \Sigma} q(x)\epsilon_x^2},$$

where, $q(x) = \Pr_{\mathcal{U}}(x)$. So $q(x) = 2^q$. From Claim A.6 we have that $\epsilon_x = 10\epsilon q^2/n$. So we have

$$d(\mu^q, \mathcal{U}) \leq 2\sqrt{\sum_{i=1}^{n} (10\epsilon q^2/n)^2}$$

Thus if $q \leq \sqrt[4]{n}/20\sqrt{\epsilon}$ we have $d(\mu^q, \mathcal{U}) \leq 2\sqrt{1/40}$ which is less than $1/3$

A.2.1 Proof of Claim A.5. Let $p(x) = \Pr_P(x)$ and $q(x) = \Pr_Q(x)$. By definition

$$H(P,Q)^2 = \frac{1}{2} \sum_{x \in \Sigma} \left(\sqrt{p(x)} - \sqrt{q(x)}\right)^2 = \left(1 - \sum_{x \in \Sigma} \sqrt{p(x)q(x)}\right)$$

Now $\sqrt{p(x)q(x)} = q(x)\sqrt{1+\epsilon_x}$. Now it is easy to verify that for all x such that $|x| \leq 1$ we have

$$\sqrt{1+x} \geq 1 + \frac{x}{2} - \frac{x^2}{2}$$

So, from the above observation we have,

$$\sqrt{p(x)q(x)} = q(x)\sqrt{1+\epsilon_x} \geq q(x)\left(1 + \frac{\epsilon_x}{2} - \frac{\epsilon_x^2}{2}\right)$$

Now since $\sum_x q(x) = 1$ and $\sum_x q_x\epsilon_x = 0$ so we have

$$H(P,Q)^2 \leq \left(1 - \sum_x q(x)\left(1 + \frac{\epsilon_x}{2} - \frac{\epsilon_x^2}{2}\right)\right) = \frac{1}{2}\sum_{x \in \Sigma} q(x)\epsilon_x^2$$

A.2.2 Proof of Claim A.6. Let $x \in \Sigma$ has k 1's and $(q-k)$ 0's. Since the μ_i is either the distribution D_1 with probability $1/2$ or distribution D_2 with probability $1/2$, so probability of x appearing when drawn from μ_i^q is

$$\frac{1}{2}\left((\frac{1}{2}+\epsilon')^k(\frac{1}{2}-\epsilon')^{q-k} + (\frac{1}{2}-\epsilon')^k(\frac{1}{2}+\epsilon')^{q-k}\right)$$
$$= \frac{1}{2^q}\frac{(1+2\epsilon')^k(1-2\epsilon')^{q-k} + (1-2\epsilon')^k(1+2\epsilon')^{q-k}}{2}$$

Using the inequality $(1+x)^r \geq 1 + xr$ (holds for $x \geq -1$ and $r \in \mathbb{N}$), we have

$$(1+2\epsilon')^k(1-2\epsilon')^{n-k} + (1-2\epsilon')^k(1+2\epsilon')^{n-k} \geq (1+2k\epsilon')(1-2(q-k)\epsilon') + (1-2k\epsilon')(1+2(q-k)\epsilon')$$

The right hand side of the above inequality is equal to $(2 - 8k(q-k)\epsilon'^2)$. Thus we have

$$\Pr_{\mu_i^q}(x) \geq \frac{1}{2^q}(1 - 4k(q-k)\epsilon'^2) \geq \frac{1}{2^q}\left(1 - \frac{4q^2\epsilon}{n}\right)$$

For the upper bound we shall use the following inequality. Let $r \in \mathbb{N}, x \geq -1$ be such that $xr < 1$. It holds that

$$(1+x)^r \leq 1 + xr + x^2r^2$$

The above inequality can be easily proved using the following facts.

(1) **When** $r \in \mathbb{N}, x > 0$ **and** $rx < 1$

(a) it holds that $(1+x)^r \leq e^{rx}$.
(b) as $0 \leq rx < 1$ it holds that $e^{rx} \leq 1 + xr + x^2r^2$.

(2) **When** $r \in \mathbb{N}, -1 \leq x \leq 0$ it holds that $(1+x)^r \leq 1 + xr + x^2r^2$ (can be proved using induction on r).

Since $\epsilon' = 2\sqrt{\epsilon/n}$ and $q \leq \sqrt[4]{n}$, $\epsilon' q < 1$. Hence, for all $k \leq q$,

$$(1+2\epsilon')^k(1-2\epsilon')^{q-k} \leq (1 + 2k\epsilon' + 4k^2\epsilon'^2)(1 - 2(q-k)\epsilon' + 4(q-k)^2\epsilon'^2),$$
$$(1-2\epsilon')^k(1+2\epsilon')^{q-k} \leq (1 - 2k\epsilon' + 4k^2\epsilon'^2)(1 + 2(q-k)\epsilon' + 4(q-k)^2\epsilon'^2)$$

and thus

$$\frac{(1+2\epsilon')^k(1-2\epsilon')^{q-k} + (1-2\epsilon')^k(1+2\epsilon')^{q-k}}{2} \leq \left(1 + 2\epsilon'^2q^2 + q^4\epsilon'^4\right)$$

Since $\epsilon' = 2\sqrt{\epsilon/n}$ and $q \leq \sqrt[4]{n}$ so we have

$$\left(1 + 2\epsilon'^2q^2 + q^4\epsilon'^4\right) \leq \left(1 + \frac{10\epsilon q^2}{n}\right)$$

And thus we have

$$\frac{1}{2^q}\left(1 - \frac{4q^2\epsilon}{n}\right) \leq \Pr_{x \leftarrow \mu_i^q}(x) \leq \frac{1}{2^q}\left(1 + \frac{10\epsilon q^2}{n}\right)$$