Karp Reduction From Entropy Gap To Statistical Difference

Ari*

June 16, 2024

Abstract

These notes are based on [Vadhan, 1999, Section 3.4].

1 Notation And Basic Definitions

For any probability distribution $\mathcal{D}$, we use $x \leftarrow\!\$\ \mathcal{D}$ to denote the event that x was sampled according to distribution $\mathcal{D}$. For any set $\mathcal{Y}$, we use notation $x \leftarrow\!\$\ \mathcal{Y}$ to denote x was sampled uniformly at random from the set $\mathcal{Y}$. We denote with $\Delta(\mathcal{X})$ the set of all probability distributions over some set $\mathcal{X}$. For any $\mathcal{D} \in \Delta(\mathcal{X})$, we use $\mathcal{D}(x)$ to denote the probability of seeing x when sampling according to distribution $\mathcal{D}$ i.e $\mathcal{D}(x) = \Pr_{x \leftarrow\!\$\ \mathcal{D}}[x]$. Similarly, given a set A, we use $\mathcal{D}(A)$ to denote $\Pr_{x \leftarrow\!\$\ \mathcal{D}}[x \in A]$. For any distribution $\mathcal{D}$, we use $\mathsf{Supp}(\mathcal{D})$ to denote the support of $\mathcal{D}$. Unless specified otherwise, we use $\log(\cdot)$ as short hand for $\log_2(\cdot)$.

Definition 1.1 (Entropy)**.** If $\mathcal{D}$ is a discrete probability distribution over some domain $\mathcal{X}$, then the entropy of $\mathcal{D}$, denoted by $\mathsf{H}(\mathcal{D})$, is defined as

$$\mathsf{H}(\mathcal{D}) = \sum_x \mathcal{D}(x) \cdot \log \frac{1}{\mathcal{D}(x)} = \mathbb{E}_{x \leftarrow\!\$\ \mathcal{D}}\left[\log \frac{1}{\mathcal{D}(x)}\right]$$

Definition 1.2 (Distributions Encoded By Circuits)**.** Let $\mathcal{D}$ be a boolean circuit (with possibly unbounded fan in) with m input gates and n output gates. The distribution encoded by $\mathcal{D}$ is a probability distribution on the set $\{0,1\}^n$, induced by feeding $\mathcal{D}$ with inputs sampled uniformly randomly from $\{0,1\}^m$. We abuse notation by referring to $\mathcal{D}$ directly as the distribution instead of the circuit for the remainder of this document. When we sample from distributions this way – we will often say that we have **white-box** access to the distribution.

2 High Level Goal/Problem Statement

We are given sample access to two distributions induced by circuits (Definition 1.2) $\mathcal{D}_1$ and $\mathcal{D}_2$ with m input gates and n output gates. We are *promised* that the two distributions have an entropy gap or entropy difference i.e we are guaranteed that $|\mathsf{H}(\mathcal{D}_1) - \mathsf{H}(\mathcal{D}_2)| \geq 1$. More formally,

*University Of Warwick

Definition 2.1 (Entropy Difference)**.** Entropy difference is the promise problem $\mathsf{ED} = \{(\mathsf{ED}^{(n)}_{\mathsf{Yes}}, \mathsf{ED}^{(n)}_{\mathsf{No}})\}_{n\in\mathbb{N}}$ where

$$\mathsf{ED}^{(n)}_{\mathsf{Yes}} = \{\mathcal{D}_1, \mathcal{D}_2 : \mathsf{H}(\mathcal{D}_1) \geq \mathsf{H}(\mathcal{D}_2) + 1\}$$
$$\mathsf{ED}^{(n)}_{\mathsf{No}} = \{\mathcal{D}_1, \mathcal{D}_2 : \mathsf{H}(\mathcal{D}_2) \geq \mathsf{H}(\mathcal{D}_1) + 1\}$$

Here $\mathcal{D}_1, \mathcal{D}_2$ are probability distributions over $\mathcal{Y} = \{0,1\}^n$.

Problem 1. We want to show that the Entropy Difference promise problem karp reduces to the Statistical Difference promise problem described below. Throughout this document when we say polynomial time, we imply polynomial time in the size of the domain of the distribution, which using the above notation is n. So we have access to two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$, which have an entropy gap. We want to show that there exists some algorithm A runs[a] in time $\mathsf{poly}(n)$ to produce to different distributions $\mathcal{D}_A$ and $\mathcal{D}_B$ such that they are either promised to at most β apart or at least α apart in variation distance.

[a] sampling from a distribution is considered to be 1 time step

Definition 2.2 (Statistical Difference)**.** For constants $0 \leq \beta < \alpha \leq 1$ such that $\alpha^2 > \beta$, Statistical Difference is a the promise problem $(\mathsf{SD}^{\alpha,\beta}) = \{\mathsf{SD}^{\alpha,\beta,n}_{\mathsf{Yes}}, \mathsf{SD}^{\alpha,\beta,n}_{\mathsf{No}}\}_{n\in\mathbb{N}}$, where

$$\mathsf{SD}^{\alpha,\beta,n}_{\mathsf{Yes}} = \{(\mathcal{D}_1, \mathcal{D}_2) : d_{\mathsf{TV}}(\mathcal{D}_1, \mathcal{D}_2) > \alpha\}$$
$$\mathsf{SD}^{\alpha,\beta,n}_{\mathsf{No}} = \{(\mathcal{D}_1, \mathcal{D}_2) : d_{\mathsf{TV}}(\mathcal{D}_1, \mathcal{D}_2) < \beta\}$$

where $\mathcal{D}_1$ and $\mathcal{D}_2$ are distributions encoded by circuits with n output gates.

3 Important Lemmas

Attention Clément: These leftover hash lemmas are originally from [Impagliazzo et al., 1989] but there are good notes for these `https://www.cs.bu.edu/~reyzin/teaching/s11cs937/notes-leo-1.pdf`. My conjecture is that if we can get left over hashing lemmas for $||\mathcal{D}||_3^3$ instead of $\mathsf{H}(\mathcal{D})$, the rest will follow easily.

Definition 3.1 (Flat Distributions)**.** A distribution $\mathcal{D}$ is flat if it is uniformly distributed over its support. If $\mathcal{D}$ is a flat distribution then, we have $|\mathsf{Supp}(\mathcal{D})| = 2^{\mathsf{H}(\mathcal{D})}$. If the support size is equal to the size of the domain, then the corresponding flat distribution is the uniform distribution.

Definition 3.2 (Universal Hash Functions)**.** A set $\mathcal{H}$ of functions mapping a domain $\mathcal{X}$ to a range $\mathcal{Y}$ is 2-universal if for every $x \neq y \in \mathcal{X}$, and $a, b \in \mathcal{Y}$,

$$\Pr_{h\xleftarrow{\$}\mathcal{H}}[h(x) = a \wedge h(b) = b] = \frac{1}{|\mathcal{Y}|^2}$$

We denote with $\mathcal{H}_{m,n}$ any 2-universal set of functions from mapping $\mathcal{X} = \{0,1\}^m$ to $\mathcal{Y} = \{0,1\}^n$.

Lemma 3.1: Leftover Hash Lemma For Flat Distributions with High Entropy

Let $\mathcal{H}$ be a 2-universal set of hash functions mapping a domain $\mathcal{X}$ to a range $\mathcal{Y}$. Let $\mathcal{D} \in \Delta(\mathcal{X})$ be a

flat distribution such that $|\mathcal{Y}| \leq \varepsilon \cdot 2^{\mathsf{H}(\mathcal{D})} = \varepsilon \cdot |\mathsf{Supp}(\mathcal{D})|$.

1. Let $\mathcal{D}_A \in \Delta(\mathcal{H} \times \mathcal{Y})$ denote the distribution induced by the following process – Pick $h \leftarrow\$ \mathcal{H}$, and then sample $x \leftarrow\$ \mathcal{D}$. Finally output $(h, h(x))$.

$$\mathcal{D}_A(h, y) = \Pr_{h \leftarrow\$ \mathcal{H}, x \leftarrow\$ \mathcal{D}} [(h, h(x) = y)]$$

2. Let $\mathcal{D}_B \in \Delta(\mathcal{H} \times \mathcal{Y})$ denote the uniform distribution over $\mathcal{H} \times \mathcal{Y}$.

$$\mathcal{D}_B(h, y) = \Pr_{h \leftarrow\$ \mathcal{H}, y \leftarrow\$ \mathcal{Y}} [(h, y)]$$

Then we have

$$d_{\mathsf{TV}}(\mathcal{D}_A, \mathcal{D}_B) \leq \varepsilon^{\Omega(1)} \tag{1}$$

On the contrary, if the *flat* distribution we are sampling from has small support (or low entropy), then sampling a hash function uniformly and then hashing samples from the distribution produces a distribution far away from uniform.

Lemma 3.2: Leftover Hash Lemma For Flat Distributions with Low Entropy

Let $\mathcal{H}$ be a 2-universal family of hash functions mapping a domain $\mathcal{X}$ to a range $\mathcal{Y}$. Let $\mathcal{D} \in \Delta(\mathcal{X})$ be a flat distribution such that $|\mathsf{Supp}(\mathcal{D})| = 2^{\mathsf{H}(\mathcal{D})} \leq \varepsilon \cdot |\mathcal{Y}|$.

1. Let $\mathcal{D}_A$ denote the distribution induced by the following process – Pick $h \leftarrow\$ \mathcal{H}$, and then sample $x \leftarrow\$ \mathcal{D}$. Finally output $(h, h(x))$.

$$\mathcal{D}_A(h, y) = \Pr_{h \leftarrow\$ \mathcal{H}, x \leftarrow\$ \mathcal{D}} [(h, h(x) = y)]$$

2. Let $\mathcal{D}_B$ denote the uniform distribution over $\mathcal{H} \times \mathcal{Y}$.

$$\mathcal{D}_B(h, y) = \Pr_{h \leftarrow\$ \mathcal{H}, y \leftarrow\$ \mathcal{Y}} [(h, y)]$$

Then we have

$$d_{\mathsf{TV}}((\mathcal{D}_A, \mathcal{D}_B)) \geq 1 - \varepsilon \tag{2}$$

Lemma 3.3: Generalised Left Over Hash Lemma

Let $\mathcal{H}$ be a family of 2-universal hash functions mapping a domain $\mathcal{X}$ to a range $\mathcal{Y}$. Let $\mathcal{D} \in \Delta(\mathcal{X})$ such that least $1 - \delta$ fraction of the support elements have probability mass at most $\frac{\varepsilon}{|\mathcal{Y}|}$.

1. $\mathcal{D}_A$: The induced distribution from the process of sampling $h \leftarrow\$ \mathcal{H}$ and $x \leftarrow\$ \mathcal{D}$, and outputting $(h, h(x)) \leftarrow\$ \mathcal{D}_A$).
2. $\mathcal{D}_B = \mathsf{Uniform}[\mathcal{H} \times \mathcal{Y}]$.

Then

$$d_{\mathsf{TV}}(\mathcal{D}_A, \mathcal{D}_B) \leq O(\delta + \varepsilon^{1/3})$$

4 Simpler Problem

We begin with two simplifying assumptions

1. We will assume that the entropy gap is large i.e for some large $k = O(n)$. We are promised that $|\mathsf{H}(\mathcal{D}_1) - \mathsf{H}(\mathcal{D}_2)| \geq k$.
2. Assume that apart from having an entropy gap we are *also* promised that both distributions are *flat* (Definition 3.1) .

4.1 An Inefficient Reduction

We first start with a reduction that is not polynomial time in n the size of the domain of the distributions. Then we proceed to make the reduction efficient by using the whiteboxness of the distribution for which we have sampling access. To reduction is as follows:

Algorithm 1: Reduction For Flat Distributions With Large Entropy Gap

1. Set $\mathcal{D} = \mathcal{D}_2$ in the above lemmas (Lemma 3.2 and Lemma 3.1). This is the distribution from which we sample before applying a randomly sampled hash function $h \xleftarrow{\$} \mathcal{H}$.
2. Then pick $\mathcal{Y} \subseteq \{0,1\}^n$ such that $|\mathcal{Y}| = 2^{\mathsf{H}(\mathcal{D}_1)} = \mathsf{Supp}(\mathcal{D}_1)$.

The distributions $\mathcal{D}_A$ and $\mathcal{D}_B$ in the lemmas above can be computed efficiently. Note that if $\mathsf{H}(\mathcal{D}_1) - \mathsf{H}(\mathcal{D}_2) \geq k$, it implies that $|\mathsf{Supp}(\mathcal{D}_1)| - |\mathsf{Supp}(\mathcal{D}_1)| \geq k$ and we are in the setting of Lemma 3.2. This would imply $d_{\mathsf{TV}}(\mathcal{D}_A, \mathcal{D}_B) \leq 2^{-k}$. If $\mathsf{H}(\mathcal{D}_1) \gg \mathsf{H}(\mathcal{D}_2)$, it implies that $|\mathsf{Supp}(\mathcal{D}_1)| \gg |\mathsf{Supp}(\mathcal{D}_2)|$, and thus, we are in the setting of Lemma 3.1. This would imply $d_{\mathsf{TV}}(\mathcal{D}_A, \mathcal{D}_B) \geq 1 - 2^{-k}$. The only thing left to show is that if we can efficiently do steps (1) and (2) described above, then we are DONE! .

Step 1, is easy as we just need to sample $h \xleftarrow{\$} \mathcal{H}$ which we know how to do efficiently. Unfortunately, Step 2, is hard. We do not know any efficient algorithm that can estimate $\mathsf{H}(\mathcal{D})$ from just sample access to distributions. If we could then, the entropy gap problem would be in BPP and there would be no need of a prover. Finding an efficient solution to estimate $\mathsf{H}(\mathcal{D})$ would imply a collapse of the polynomial time hierarchy. Luckily, we have a little more than just sample access to distributions $\mathcal{D}_1$ and $\mathcal{D}_2$. We have access to circuits that induce these distributions.

Attention Clément: In [Herman and Rothblum, 2022] they do not have this issue. $\mathcal{D}_1$ is given to the verifier by the prover in the form of a histogram. The verifier can set $\mathcal{Y} = 2^{\mathsf{H}(\mathcal{D}_1)}$ by just computing $\mathsf{H}(\mathcal{D}_1)$. The distribution $\mathcal{D}_2$ is the one the verifier has sample access to. So they do not need this circuit trick from [Okamoto, 1996] that Salil uses in his thesis that is described below.

4.2 Getting around estimating entropy

In this section, for a distribution $\mathcal{D}$, we use $C_{\mathcal{D}} : \{0,1\}^m \to \{0,1\}^n$ to denote the circuit that induces this distribution. Define $\Omega_{\mathcal{D}_1}(x) = \{r \in \{0,1\}^m : C_{\mathcal{D}_1}(r) = x\}$. Since $\mathcal{D}_1$ is flat, if $x \in \mathsf{Supp}(\mathcal{D}_1)$, then we have

$$|\Omega_{\mathcal{D}_1}(x)| = 2^m \cdot \mathcal{D}_1(x) = 2^m \cdot 2^{-\mathsf{H}(\mathcal{D}_1)}$$

For $x \notin \mathsf{Supp}(\mathcal{D}_1)$, we have

$$|\Omega_{\mathcal{D}_1}(x)| = 2^m \cdot \mathcal{D}_1(x) = 0$$

Thus for any $x \in \mathsf{Supp}(\mathcal{D}_1)$, we have

$$|\mathsf{Supp}(\mathcal{D}_2) \times \Omega_{\mathcal{D}_1}(x)| = 2^{\mathsf{H}(\mathcal{D}_2) - \mathsf{H}(\mathcal{D}_1) + m}$$

Thus $\mathsf{H}(\mathcal{D}_2) \gg \mathsf{H}(\mathcal{D}_1) \implies |S_{\mathcal{D}_2} \times \Omega_{\mathcal{D}_1}(x)| \gg 2^m$ and $\mathsf{H}(\mathcal{D}_1) \gg \mathsf{H}(\mathcal{D}_2) \implies |S_{\mathcal{D}_2} \times \Omega_{\mathcal{D}_1}(x)| \ll 2^m$, where m is available to us as we know the circuits that encode the distributions. So instead of hashing samples

from $\mathcal{D}_2$ to $\mathcal{Y} = 2^{\mathsf{H}(\mathcal{D}_1)}$ as specified by the algorithm, we will hash samples from the uniform distribution on $\mathsf{Supp}(\mathcal{D}_2) \times \Omega_{\mathcal{D}_1}(x)$ for some $x \in \mathsf{Supp}(\mathcal{D}_1)$. In other words, in the lemmas above we set $\mathcal{Y} = \{0,1\}^m$, and $\mathcal{D} = \mathsf{Uniform}\left[\mathsf{Supp}(\mathcal{D}_2) \times \Omega_{\mathcal{D}_1}(x)\right]$. Accordingly, define $\mathcal{H} = \mathcal{H}_{m+n,m}$, and let $\mathcal{D}_A$ and $\mathcal{D}_B$ as the following distributions (note we are still doing the same thing as we did earlier just with distributions with slightly larger domains).

$\mathcal{D}_A$: Choose $r \xleftarrow{\$} \{0,1\}^m$ and let $x = \mathcal{D}_1(r)$. Choose $h \xleftarrow{\$} \mathcal{H}$ and $y \xleftarrow{\$} \mathcal{D}_2$. Output $(x, h, h(r||y))$. This is equivalent to saying $(x, h, h(r, y)) \xleftarrow{\$} \mathcal{D}_A$

$\mathcal{D}_B$: Choose $x \xleftarrow{\$} \mathcal{D}_1$, $h \xleftarrow{\$} \mathcal{H}$ and $z \xleftarrow{\$} \{0,1\}^m$. Output (x, h, z). This is equivalent to saying $(x, h, z) \xleftarrow{\$} \mathcal{D}_B$

Thus by Lemma 3.2

$$\mathsf{H}(\mathcal{D}_1) > \mathsf{H}(\mathcal{D}_2) + k \implies d_{\mathsf{TV}}(\mathcal{D}_A, \mathcal{D}_B) \geq 1 - 2^{-\Omega(k)}$$

By Lemma 3.1,

$$\mathsf{H}(\mathcal{D}_1) < \mathsf{H}(\mathcal{D}_2) + k \implies d_{\mathsf{TV}}(\mathcal{D}_A, \mathcal{D}_B) \leq 2^{-\Omega(k)}$$

But we still have not answered the question of how we efficiently sample from $\Omega_{\mathcal{D}_1}(x)$? We simply simulate it by sampling $r \xleftarrow{\$} \{0,1\}^m$, and compute $x = C_{\mathcal{D}_1}(r)$. Now conditioned on x, r was uniformly sampled from $\Omega_{\mathcal{D}_1}(x)$. This gives us a *single* sample $r \xleftarrow{\$} \Omega_{\mathcal{D}_1}(x)$.

At this point we have solved our problem, given our assumptions (1) the entropy gap was large enough, and (2) the distributions were *flat*.

5 Removing Assumptions

5.1 Small Entropy Gap To Large Entropy Gap

What happens if we do *not* have that $|\mathsf{H}(\mathcal{D}_1) - \mathsf{H}(\mathcal{D}_2)| \geq k$. The solution is simple, just replace $\mathcal{D}_1$ and $\mathcal{D}_2$ above with k independent copies i.e. $\otimes^k \mathcal{D}_1$ and $\otimes^k \mathcal{D}_2$. As we have $\mathsf{H}(\otimes^k \mathcal{D}_1) = k\mathsf{H}(\mathcal{D}_1)$ and $\mathsf{H}(\otimes^k \mathcal{D}_2) = k\mathsf{H}(\mathcal{D}_2)$. We get $|\mathsf{H}(\otimes^k \mathcal{D}_1) - \mathsf{H}(\otimes^k \mathcal{D}_2)| \geq k$, as we are promised that $|\mathsf{H}(\mathcal{D}_1) - \mathsf{H}(\mathcal{D}_2)| \geq 1$.

5.2 Non-Flat Distributions

So the only question we need to resolve is what happens when $\mathcal{D}_1$ and $\mathcal{D}_2$ are not flat ? It turns out that even if $\mathcal{D}_1$ and $\mathcal{D}_2$ are not flat, and if we sample "enough" independent copies, they become "nearly" flat. The task of this section is to define precisely, how many copies is enough, and what "nearly" means.

Definition 5.1 (Heavy, Light And Typical Elements Of A Distribution)**.** Let $\mathcal{D} \in \Delta(\mathcal{X})$ and fix $\eta \in \mathbb{R}^+$. For any $x \in \mathsf{Supp}(\mathcal{D})$

1. x is η-heavy if $\mathcal{D}(x) \geq 2^{\eta - \mathsf{H}(\mathcal{D})}$.
2. x is η-light if $\mathcal{D}(x) \leq 2^{-\eta - \mathsf{H}(\mathcal{D})}$
3. x is η-typical otherwise.

Definition 5.2 (Nearly Flat Distributions)**.** Fix $\beta \in \mathbb{R}^+$ and pick an arbitrary $c > 0$. For a distribution $\mathcal{D}$, let $T^{(\mathcal{D})}_{c\beta} = \{x \in \mathsf{Supp}(\mathcal{D}) : x \text{ is } c\beta - \text{typical}\}$ denote the set of $c\beta$ typical elements in the support of $\mathcal{D}$. We say $\mathcal{D}$ is β-flat if for every $c > 0$,

$$\mathcal{D}(T^{(\mathcal{D})}_{c\beta}) = \Pr_{x \leftarrow\$ \mathcal{D}}\left[x \in T^{(\mathcal{D})}_{c\beta}\right] \geq 1 - 2^{-c^2+1}$$

The above definition can be thought of as saying that a "nearly flat" distribution has all but negligible support being non-typical (not that many light or heavy elements).

Lemma 5.1: Repetition flattens a distribution

Let $\mathcal{D}$ be a distribution induced by a circuit with m inputs and k be a positive integer, and $\otimes^k \mathcal{D}$ denote k independent copies of $\mathcal{D}$. Suppose[a] for all $x \in \mathsf{Supp}(\mathcal{D})$, we have $\Pr_{X \leftarrow\$ \mathcal{D}}[X = x] \geq \frac{1}{2^m}$. Then $\otimes^k \mathcal{D}$ is $\sqrt{k} \cdot m$-flat.

[a]Which the circuit representation of the circuit guarantees us. There has to be one coin $r \in \{0,1\}^m$ such that the circuit outputs x for x to be in $S_{\mathcal{D}}$.

Attention Clément: In [Herman and Rothblum, 2022] they assume that they can ignore elements with very small probability. So they do not really need the circuit induced distributions to get for all $x \in \mathsf{Supp}(\mathcal{D})$ $\mathcal{D}(x) \geq \frac{1}{2^m}$.

Proof. Fix any $c > 0$ and $\beta \in \mathbb{R}^+$ whose value will defined later. For any $x \in \mathsf{Supp}(\mathcal{D})$, Define $\mathsf{weight}(x) = -\log(\mathcal{D}(x))$. Thus $\mathsf{weight}(x) \in [0, m]$[1]. For $\vec{x} \leftarrow\$ \otimes^k \mathcal{D}$ we say $\vec{x}$ is $c\beta$-typical if for all $i \in [k]$, x_i is $c\beta$-typical. Or in other words, we have, for $\vec{x} = (x_1, \ldots, x_k)$, for all $i \in [k]$,

$$-c\beta \leq wt(x_i) - \mathsf{H}(\mathcal{D}) \leq c\beta \tag{3}$$

Thus

$$\Pr_{\vec{x} \leftarrow\$ \otimes^k \mathcal{D}}[\vec{x} \text{ is } \textit{not} \text{ atypical }] = \Pr_{\vec{x} \leftarrow\$ \otimes^k \mathcal{D}}[\exists x_j : x_j \text{ is } \textit{not} \text{ atypical }] \tag{4}$$

$$\leq \Pr_{\vec{x} \leftarrow\$ \otimes^k \mathcal{D}}\left[\left|\frac{1}{k}\sum_{i \in [k]} \mathsf{weight}(x_i) - \mathsf{H}(\mathcal{D})\right| \geq \frac{c\beta}{k}\right] \tag{5}$$

Equation (5) comes from the union bound. Note $\mathsf{weight}(x_i)$ is a bounded random variable in $[0, m]$ and $\mathbb{E}_{x_i \leftarrow\$ \mathcal{D}}[\mathsf{weight}(x_i)] = \mathsf{H}(\mathcal{D})$. Thus, using Hoeffding Inequality, we get

$$\Pr_{\vec{X} \leftarrow\$ \otimes^k \mathcal{D}}\left[\left|\frac{1}{k}\sum_{i \in [k]} wt(X_i) - \mathsf{H}(\mathcal{D})\right| \geq \frac{c}{k}\beta\right] \leq 2 \cdot \exp\left(\frac{-2 \cdot k(c\beta/k)^2}{m^2}\right) \tag{6}$$

Plugging $\beta = \sqrt{k}m$ gives us

$$\Pr_{\vec{x} \leftarrow\$ \otimes^k \mathcal{D}}\left[\left|\frac{1}{k}\sum_{i \in [k]} \mathsf{weight}(x_i) - \mathsf{H}(\mathcal{D})\right| \geq \frac{c}{k}\beta\right] \leq 2^{-c^2+1} \tag{7}$$

□

[1]By assumption $\mathcal{D}(x) \geq 2^{-m}$

6 General Construction Without Assumptions

With the above definitions in place, we are ready to derive the general construction. Let k be a large constant whose value can be thought of as the security parameter. Let $q = 9km^2$ and define $\widetilde{\mathcal{D}_1} = \otimes^q \mathcal{D}_1$ and $\widetilde{\mathcal{D}_2} = \otimes^q \mathcal{D}_2$, where $\mathcal{D}_1$ and $\mathcal{D}_2$ are distributions induced by circuits with m inputs and n outputs.

> We will play the same game as in Section 4.2, but instead of $\mathcal{D}_1$ and $\mathcal{D}_2$, we will use $\widetilde{\mathcal{D}_1}$ instead of $\mathcal{D}_1$ and $\widetilde{\mathcal{D}_2}$ for $\mathcal{D}_2$.

Let $m' = qm$ and $n' = qn$. Let $\mathcal{H} = \mathcal{H}_{m'+n',m'}$.

> 1. $\mathcal{D}_A$: The induced distribution of the following procedure: Sample $r \leftarrow\!\$\ \{0,1\}^{m'}$, and let $x = C_{\widetilde{\mathcal{D}_1}}(r)$, Sample $y \leftarrow\!\$\ \mathcal{D}_2$, Sample $h \leftarrow\!\$\ \mathcal{H}$. Output $(x, h, h(r||y))$
> 2. $\mathcal{D}_B$: The distribution induced by $x \leftarrow\!\$\ \widetilde{\mathcal{D}_1}$, $h \leftarrow\!\$\ \mathcal{H}$, and $r' \leftarrow\!\$\ \{0,1\}^{m'}$.

To make the proof more readable, we split each component of the above distributions into 3 parts. Let $\mathcal{D}_A = (\mathcal{D}_{A_1}, \mathcal{D}_{A_2}, \mathcal{D}_{A_3})$ and $\mathcal{D}_B = (\mathcal{D}_{B_1}, \mathcal{D}_{B_1}, \mathcal{D}_{B_3})$ be those parts. To show $\mathcal{D}_A$ and $\mathcal{D}_B$ are statistically far, it suffices to show, $\mathcal{D}_{A_3}$ and $\mathcal{D}_{B_3}$ are far, conditioned on the event $x \leftarrow\!\$\ \widetilde{\mathcal{D}_1}$ and $h \leftarrow\!\$\ \mathcal{H}$. By Lemma 5.1, $\widetilde{\mathcal{D}_1}$ and $\widetilde{\mathcal{D}_2}$ are $\beta = \sqrt{q}m = \sqrt{9km^2} \cdot m = 3\sqrt{k}m^2$ flat. Since $\widetilde{\mathcal{D}_1}$ is $3\sqrt{k}m^2$ flat, setting $c = \sqrt{k}$, we have $\widetilde{\mathcal{D}_1}(x \text{ is } \beta\sqrt{k} \text{ typical}) \geq 1 - 2^{-k+1}$.

Remark. *When we say x is typical, unless specified otherwise, we mean x is $\beta\sqrt{k}$-typical. To make things easier to read, we will highlight x in blue, to denote that we are dealing with a typical x.*

We write $\mathcal{D}_{A_{x,h}}$ as the distribution $\mathcal{D}_{A_3}$ conditioned on the event that $x \leftarrow\!\$\ \widetilde{\mathcal{D}_1}$ is typical. Similarly, we define $\mathcal{D}_{B_{x,h}}$ as the distribution $\mathcal{D}_{B_3}$ conditioned on x being typical. From the definition of $\mathcal{D}_{B_3}$, we have $\mathcal{D}_{B_{x,h}}$ is distributed uniformly over $\{0,1\}^{m'}$ (there is no difference between $\mathcal{D}_{B_3}$ and $\mathcal{D}_{B_{x,h}}$).

$\mathcal{D}_{A_{x,h}}$ on the other hand is different from $\mathcal{D}_{A_3}$, and it denotes the distribution induced by outputting $h(r||y)$, when $(r, y) \leftarrow\!\$\ \Omega_{\widetilde{\mathcal{D}_1}}(x) \times \widetilde{\mathcal{D}_2}$, with the important constraint that the x in $\Omega_{\widetilde{\mathcal{D}_1}}(x)$ is typical. Note as $\widetilde{\mathcal{D}_2}$ is also β-flat, we have $\widetilde{\mathcal{D}_2}(y \text{ is typical}) \geq 1 - 2^{-k+1}$ (setting $c = \sqrt{k}$). Now we define the set,

$$T_{x,h} = \{h(r, y) : r \in \Omega_{\widetilde{\mathcal{D}_1}}(x) \wedge y \text{ is typical}\}$$

We have

$$\mathcal{D}_{A_{x,h}}(T_{x,h}) \geq 1 - 2^{-k+2} \tag{8}$$

Why? The probability that we sample a non-typical $x \leftarrow\!\$\ \widetilde{\mathcal{D}_1}$ is at most 2^{-k+1}, and the the probability that we sample non typical $y \leftarrow\!\$\ \widetilde{\mathcal{D}_2}$ is at most 2^{-k+1}. $\mathcal{D}_{A_{x,h}}(T_{x,h})$ is the probability that I sample $z \leftarrow\!\$\ A_{x,h}$ and $z \in T_{x,h}$. For z to be in $T_{x,h}$, two things have to happen, when I sample $x \leftarrow\!\$\ \widetilde{\mathcal{D}_1}$, x must be typical, and when I sample $y \leftarrow\!\$\ \widetilde{\mathcal{D}_2}$ to feed into h, I need y to be typical. By the union bound we get what we want.

Note that $|T_{x,h}| \leq |\Omega_{\widetilde{\mathcal{D}_1}}(x)| \cdot n_{\texttt{typical}}$, where $n_{\texttt{typical}}$ denotes the number of $\sqrt{k}\beta$ typical y's in $\mathsf{Supp}(\widetilde{\mathcal{D}_2})$. Finally we get,

$$|\Omega_{\widetilde{\mathcal{D}_1}}(x)| = 2^{m'} \cdot \widetilde{\mathcal{D}_1}(x \text{ is typical}) \tag{9}$$

$$\leq 2^{m'} 2^{\sqrt{k}\beta - \mathsf{H}(\widetilde{\mathcal{D}_1})} \tag{10}$$

and

$$n_{\texttt{typical}} \leq 2^{\sqrt{k}\beta + \mathsf{H}(\widetilde{\mathcal{D}_2})} \tag{11}$$

For (9) there are $2^{m'}$ values for r, but the number of r's that map to any x depends on the probability of seeing that x. (10) comes from that definition of typical (Defintion 5.1) which says for x to be η-typical, we must have $\widetilde{\mathcal{D}_1}(x) \leq 2^{-\eta-\mathsf{H}(\widetilde{\mathcal{D}_1})}$. So that's the largest $\widetilde{\mathcal{D}_1}(x)$ can be. Here $\eta = \sqrt{k}\beta$. (11) since each typical y must have mass at least $2^{-\mathsf{H}(\widetilde{\mathcal{D}_2})-\sqrt{k}\beta}$ (Defintion 5.1). Also[2] note that $q > 2\sqrt{k}\beta + k$. Now we have all the machinery to prove the final statement.

Yes CASE : Now assume that $(\mathcal{D}_1, \mathcal{D}_2) \in \mathsf{ED}_{\mathsf{Yes}}$ i.e. $\mathsf{H}(\mathcal{D}_1) > \mathsf{H}(\mathcal{D}_2) + 1$. Then $\mathsf{H}(\widetilde{\mathcal{D}_1}) > \mathsf{H}(\widetilde{\mathcal{D}_2}) + q$, where $q > 2\sqrt{k}\beta + k$ (as repetition amplifies the gap).

$$
\begin{aligned}
|T_{x,h}| &\leq n_{\mathtt{typical}} \cdot |\Omega_{\widetilde{\mathcal{D}_1}}(x)| && (12)\\
&\leq 2^{\sqrt{k}\beta+\mathsf{H}(\widetilde{\mathcal{D}_2})} 2^{m'+\sqrt{k}\beta-\mathsf{H}(\widetilde{\mathcal{D}_1})} && (13)\\
&= 2^{m'+\mathsf{H}(\widetilde{\mathcal{D}_2})-\mathsf{H}(\widetilde{\mathcal{D}_1})+2\sqrt{k}\beta} && (14)\\
&\leq 2^{m'-q+2\sqrt{k}\beta} && (15)\\
&\leq 2^{m'-k} && (16)
\end{aligned}
$$

Equations (15) and (16) comes from the fact that $q > 2\sqrt{k}\beta + k$.
Finally, from the definition of total variation difference

$$
\begin{aligned}
d_{\mathsf{TV}}\left(A_{x,h}, B_{x,h}\right) &= \max_S |\mathcal{D}_{A_{x,h}}(S) - \mathcal{D}_{B_{x,h}}(S)| && (17)\\
&\geq \mathcal{D}_{A_{x,h}}(T_{x,h}) - \mathcal{D}_{B_{x,h}}(T_{x,h}) && (18)\\
&\geq (1 - 2^{-k+2}) - \frac{|T_{x,h}|}{2^{m'}} && (19)\\
&\geq (1 - 2^{-k+2}) - 2^{-k} && (20)\\
&= 1 - O(2^{-k}) && (21)
\end{aligned}
$$

(19) comes from equation (8) and the fact that $\mathcal{D}_{B_{x,h}} = \mathcal{D}_{B_3}$

No CASE : Now assume that $(\mathcal{D}_1, \mathcal{D}_2) \in \mathsf{ED}_{\mathsf{No}}$ i.e. $\mathsf{H}(\mathcal{D}_2) > \mathsf{H}(\mathcal{D}_1) + 1$. Then $\mathsf{H}(\widetilde{\mathcal{D}_2}) > \mathsf{H}(\widetilde{\mathcal{D}_1}) + q$, where $q > 2\sqrt{k}\beta + k$ (as shown above). Note that $\Omega_{\widetilde{\mathcal{D}_1}}(x)$ is a flat distribution (as all $r \in \Omega_{\widetilde{\mathcal{D}_1}}(x)$ have the same probability of mapping to any x) and $\widetilde{\mathcal{D}_2}$ is a β-flat. Thus $\Omega_{\widetilde{\mathcal{D}_1}}(x) \times \widetilde{\mathcal{D}_2}$ is also β-flat. Conditioned on x being typical, and let $\mathcal{D}$ be shorthand for the sampling distribution $\Omega_{\widetilde{\mathcal{D}_1}}(x) \times \widetilde{\mathcal{D}_2}$. Then we have

$$
\begin{aligned}
\mathsf{H}(\mathcal{D}) &= \mathsf{H}(\Omega_{\widetilde{\mathcal{D}_1}}(x) \times \widetilde{\mathcal{D}_2}) && (22)\\
&= \mathsf{H}(\Omega_{\widetilde{\mathcal{D}_1}}(x)) + \mathsf{H}(\widetilde{\mathcal{D}_2}) && (23)\\
&= \log |\Omega_{\widetilde{\mathcal{D}_1}}(x)| + \mathsf{H}(\widetilde{\mathcal{D}_2}) && (24)\\
&\geq (m' - \mathsf{H}(\widetilde{\mathcal{D}_1}) - \sqrt{k}\beta) + (\mathsf{H}(\widetilde{\mathcal{D}_1}) + 2\sqrt{k}\beta + k) && (25)\\
&\geq m' + k + \sqrt{k}\beta && (26)
\end{aligned}
$$

Equation (25) comes $\widetilde{\mathcal{D}_1}$ being β-flat and thus for each $x \in \mathsf{Supp}(\mathcal{D}_1)$ that is typical, we have $\widetilde{\mathcal{D}_1}(x) \geq 2^{-\sqrt{k}\beta-\mathsf{H}(\widetilde{\mathcal{D}_1})}$. Therefore, $|\Omega_{\widetilde{\mathcal{D}_1}}(x)| = \widetilde{\mathcal{D}_1}(x) \cdot 2^m \geq 2^{-\sqrt{k}\beta-\mathsf{H}(\widetilde{\mathcal{D}_1})+m}$. The second summand is just saying $\mathsf{H}(\widetilde{\mathcal{D}_2}) > \mathsf{H}(\widetilde{\mathcal{D}_1}) + q$, where $q > 2\sqrt{k}\beta + k$

[2]Expand the RHS by plugging in $\beta = \sqrt{q}$, its clear the LHS is bigger.

Set $\mathcal{Y} = \{0,1\}^{m'}$ and $\delta = 2^{-k+1}$. As it's β-flat, we have with probability $1-\delta$, we sample a typical x. And as x is typical, we have $\mathcal{D}(x) \leq 2^{\sqrt{k}\beta - \mathsf{H}(\mathcal{D})} \leq 2^{-m'-k}$, where the last inequality comes from Equation (26). Setting $\varepsilon = 2^{-k}$, we get the same conditions for the generalised left over hashing lemma (Lemma 3.3) – at least $1-\delta$ of the fraction of elements in the support of $\mathcal{D}$ have mass at least $2^{-(m'+k)}$. Therefore, the statistical difference between $\mathcal{D}_A$ and $\mathcal{D}_B$ must be at most $O(2^{-k/3} + 2^{-k+1})$, which is negligible.

References

T. Herman and G. N. Rothblum. Verifying the unseen: interactive proofs for label-invariant distribution properties. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 1208–1219, 2022.

R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24, 1989.

T. Okamoto. On relationships between statistical zero-knowledge proofs. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 649–658, 1996.

S. P. Vadhan. *A study of statistical zero-knowledge proofs*. PhD thesis, Massachusetts Institute of Technology, 1999.