Characterizing the Distinguishability of Product Distributions Through Multicalibration

Cassandra Marcussen
Harvard University, Cambridge, MA, USA

Aaron Putterman
Harvard University, Cambridge, MA, USA

Salil Vadhan
Harvard University, Cambridge, MA, USA

Abstract

Given a sequence of samples $x_1, \ldots, x_k$ promised to be drawn from one of two distributions X_0, X_1, a well-studied problem in statistics is to decide *which* distribution the samples are from. Information theoretically, the maximum advantage in distinguishing the two distributions given k samples is captured by the total variation distance between $X_0^{\otimes k}$ and $X_1^{\otimes k}$. However, when we restrict our attention to *efficient distinguishers* (i.e., small circuits) of these two distributions, exactly characterizing the ability to distinguish $X_0^{\otimes k}$ and $X_1^{\otimes k}$ is more involved and less understood.

In this work, we give a general way to reduce bounds on the computational indistinguishability of X_0 and X_1 to bounds on the *information-theoretic* indistinguishability of some specific, related variables $\widetilde{X}_0$ and $\widetilde{X}_1$. As a consequence, we prove a new, tight characterization of the number of samples k needed to efficiently distinguish $X_0^{\otimes k}$ and $X_1^{\otimes k}$ with constant advantage as

$$k = \Theta\left(d_H^{-2}\left(\widetilde{X}_0, \widetilde{X}_1\right)\right),$$

which is the inverse of the squared Hellinger distance d_H between two distributions $\widetilde{X}_0$ and $\widetilde{X}_1$ that are computationally indistinguishable from X_0 and X_1. Likewise, our framework can be used to re-derive a result of Halevi and Rabin (TCC 2008) and Geier (TCC 2022), proving nearly-tight bounds on how computational indistinguishability scales with the number of samples for arbitrary product distributions.

At the heart of our work is the use of the Multicalibration Theorem (Hébert-Johnson, Kim, Reingold, Rothblum 2018) in a way inspired by recent work of Casacuberta, Dwork, and Vadhan (STOC 2024). Multicalibration allows us to relate the computational indistinguishability of X_0, X_1 to the statistical indistinguishability of $\widetilde{X}_0, \widetilde{X}_1$ (for lower bounds on k) and construct explicit circuits to distinguish between $\widetilde{X}_0, \widetilde{X}_1$ and consequently X_0, X_1 (for upper bounds on k).

2012 ACM Subject Classification Theory of computation → Circuit complexity

Keywords and phrases Multicalibration, computational distinguishability

Digital Object Identifier 10.4230/LIPIcs.CCC.2025.19

Related Version *Full Version*: https://arxiv.org/abs/2412.03562

Funding *Cassandra Marcussen*: Supported in part by an NDSEG fellowship, and by NSF Award 2152413 and a Simons Investigator Award to Madhu Sudan.
Aaron Putterman: Supported in part by the Simons Investigator Awards of Madhu Sudan and Salil Vadhan and NSF Award CCF 2152413.
Salil Vadhan: Supported by a Simons Investigator Award.

Acknowledgements This project was inspired by a final project completed by the first two authors during Cynthia Dwork's course "Topics in Theory for Society: The Theory of Algorithmic Fairness" at Harvard (Spring 2024). We thank Cynthia Dwork for initial discussions regarding multicalibration and its applications to complexity and hardness amplification. We thank the anonymous Eurocrypt and CCC reviewers for their suggestions and feedback. We thank Pranay Tankala for a correction regarding the statement of the Multicalibration Theorem.

1 Introduction

Given a sequence of samples $x_1, \ldots, x_k$ promised to be drawn from one of two distributions X_0, X_1, each defined on a domain $\mathcal{X}$, a well-studied problem in statistics is to decide *which* distribution the samples are from. Information theoretically, the maximum advantage in distinguishing X_0, X_1 given k samples is $d_{\mathrm{TV}}(X_0^{\otimes k}, X_1^{\otimes k})$, where $X_b^{\otimes k}$ denotes the result of independently sampling k times from X_b, and $d_{\mathrm{TV}}(p,q) = \frac{1}{2}\sum_{x\in\mathcal{X}} |p(x) - q(x)|$ denotes the total variation distance. This advantage can then be related to the single-sample relation between X_0, X_1 in a few ways:

1. An upper bound is given by the inequality:

$$d_{\mathrm{TV}}(X_0^{\otimes k}, X_1^{\otimes k}) \leq 1 - (1 - d_{\mathrm{TV}}(X_0, X_1))^k \leq k \cdot d_{\mathrm{TV}}(X_0, X_1). \tag{1}$$

2. To obtain a good 2-sided bound, we can use the *Hellinger distance*:

$$1 - e^{-k d_{\mathrm{H}}^2(X_0,X_1)} \leq d_{\mathrm{TV}}(X_0^{\otimes k}, X_1^{\otimes k}) \leq \sqrt{2k d_{\mathrm{H}}^2(X_0, X_1)}, \tag{2}$$

where the Hellinger distance is defined as

$$d_H^2(X_0, X_1) = \frac{1}{2}\sum_{i\in\mathcal{X}} \left(\sqrt{\Pr[X_0 = i]} - \sqrt{\Pr[X_1 = i]}\right)^2.$$

In particular, Inequality (2) above shows that $k = \Theta\left(1/d_H^2(X_0, X_1)\right)$ samples is both necessary *and* sufficient for distinguishing X_0, X_1 with constant advantage. In contrast, observe that Equation (1) in the first point above is actually *not* tight. To illustrate, we can consider the two distributions below:

1. X_0 such that $\Pr[X_0 = 1] = 1/2, \Pr[X_0 = 0] = 1/2$.
2. X_1 such that $\Pr[X_1 = 1] = 1/2 + \varepsilon, \Pr[X_1 = 0] = 1/2 - \varepsilon$.

Equation (1) implies only that $d_{\mathrm{TV}}(X_0^{\otimes k}, X_1^{\otimes k}) \leq k\varepsilon$, and thus does not rule out the possibility of distintiguishing X_0, X_1 with constant advantage after $k = O(1/\varepsilon)$ samples. However, as we see in the *Hellinger*-distance formulation of Equation (2), in this example $k = \Theta(1/\varepsilon^2)$ samples is both necessary and sufficient to distinguish X_0, X_1 with constant advantage. Thus, a key advantage in using Hellinger distance to characterize statistical distinguishability is in getting *instance-optimal* characterizations of the statistical distinguishability *for every* pair of random variables.

In computer science, this task of distinguishing distributions arises in many domains, from cryptography to computational complexity. However, we typically work with the computational analogue of total variation distance known as *computational indistinguishability*. We say that X_0, X_1 are *δ-indistinguishable by circuits of size s* if for every function $f : \mathcal{X} \to \{0,1\}$ computable by size s circuits, $|\sum_{x\in\mathcal{X}} f(x) \cdot (\Pr[X_0 = x] - \Pr[X_1 = x])| \leq \delta$. In this paper, we state most of our results using this concrete security formulation, where s and δ are separate parameters. However, our results can be translated into the traditional asymptotic complexity formulation where $s = n^{\omega(1)}$, and $\delta = n^{-\omega(1)}$ for a security parameter n and unspecified super-polynomial functions $n^{\omega(1)}$. In fact, we are interested in the setting of *weak indistinguishability* where δ is non-negligible ($\delta \geq 1/\mathrm{poly}(n)$) while s remains super-polynomial ($s \geq n^{\omega(1)}$). We elaborate on the asymptotic formulations of our results in Section 1.2.

The classic hybrid argument of Goldwasser and Micali [6] shows that if X_0, X_1 are δ-indistinguishable for circuits of size s, then $X_0^{\otimes k}, X_1^{\otimes k}$ are $k\delta$-indistinguishable for circuits of size s, which corresponds to the weaker bound in Equation (1) above. However, a more refined

understanding of (in)distinguishability under repeated samples has been elusive. Indeed, the work of Halevi and Rabin [9] was the first to show that $X_0^{\otimes k}, X_1^{\otimes k}$ are $(1-(1-\delta)^k+\varepsilon)$-indistinguishable for circuits of size $s' = s/\text{poly}(k, 1/\varepsilon)$ (with subsequent improvements to the parameters being made by [5]). Note that when $s = n^{\omega(1)}$ and $k \leq \text{poly}(n)$, we can take $\varepsilon = n^{-\omega(1)}$ and retain $s' = n^{\omega(1)}$, so this bound agrees with the tighter bound in Equation (1) above up to an additive negligible term.

This bound on computational distinguishability can be viewed as an extension of Equation (1) to the computational setting. Similarly to Equation (1) from the information-theoretic setting, the characterization of [9, 5] is *not* an instance-optimal characterization of the computational distinguishability of random variables, and a computational analog of Equation (2) is still missing. In this work, we overcome this shortcoming by building on the key technique of *multicalibration*, as we explain below.

1.1 Main Results

In this work, we give a general way to reduce reasoning about computational indistinguishability to reasoning about the information-theoretic case. This also allows us to deduce as corollaries both the characterization achieved by [9, 5] and a computational analogue of Inequality (2) (with the total variation distance of X_0, X_1 replaced by the computational indistinguishability of X_0, X_1 and the Hellinger distance between X_0, X_1 replaced by a quantity we call the *pseudo-Hellinger distance* between X_0, X_1).

▶ **Theorem 1.1.** *For every pair of random variables X_0, X_1 over $\mathcal{X}$, every integer s and every $\varepsilon > 0$, there exist random variables $\widetilde{X}_0, \widetilde{X}_1$ such that for every $k > 0$,*

1. *X_b is ε-indistinguishable from $\widetilde{X}_b$ by circuits of size s, for each $b \in \{0,1\}$.*
2. *$X_0^{\otimes k}$ and $X_1^{\otimes k}$ are $(d_{TV}(\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}) + 2k \cdot \varepsilon)$-indistinguishable by circuits of size s.*
3. *$X_0^{\otimes k}$ and $X_1^{\otimes k}$ are $(d_{TV}(\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}) - 2k \cdot \varepsilon)$-distinguishable by circuits of size $s' = O(sk/\varepsilon^6) + \text{poly}(k/\varepsilon)$.*
4. *The statements above also hold with $\widetilde{X}_1 = X_1$, but $s' = O(sk/\varepsilon^{12}) + \text{poly}(k/\varepsilon)$.*

▶ Remark 1.2. More generally, we can define indistinguishability with respect to an arbitrary class of functions $\mathcal{F}$. See the full version of the paper for generalized versions of the different parts of the above statement.

▶ Remark 1.3. We can also generalize the above theorem to arbitrary product distributions. We prove this in the full version of the paper.

As elaborated upon in Section 1.3, the above theorem follows from assigning $\widetilde{X}_0$ and $\widetilde{X}_1$ to be a careful "mixture" of X_0 and X_1. This mixture depends on carefully partitioning the domain space of the random variables using *multicalibration*. With these random variables, Item 2 follows directly from Item 1 via hybrid argument, so the value of Theorem 1.1 is that it provides not only an upper bound on (in)distinguishability but a matching lower bound, via Item 3. Thinking of $s = n^{\omega(1)}, k \leq \text{poly}(n), \varepsilon = 1/s$, we see that up to a negligible additive term ($k\varepsilon$), and a polynomial change in circuit size (s' vs. s), the computational indistinguishability of X_0, X_1 under multiple samples is tightly captured by the information-theoretic indistinguishability of $\widetilde{X}_0, \widetilde{X}_1$ under multiple samples. In particular, $k = \Theta\left(1/d_{\text{H}}^2(\widetilde{X}_0, \widetilde{X}_1)\right)$ samples are both necessary and sufficient to distinguish X_0, X_1 with constant advantage. We can abstract this latter corollary as follows:

▶ **Definition 1.4.** *For random variables X_0, X_1 over $\mathcal{X}$, $s \in \mathbb{N}$, $\varepsilon > 0$, the* (s, ε) pseudo-Hellinger distance *between X_0 and X_1 is the smallest δ such that there exist random variables $\widetilde{X}_0, \widetilde{X}_1$ such that:*

1. X_0 *is* ε*-indistinguishable from* $\widetilde{X}_0$ *for circuits of size* s.
2. X_1 *is* ε*-indistinguishable from* $\widetilde{X}_1$ *for circuits of size* s.
3. $d_{\mathrm{H}}(\widetilde{X}_0, \widetilde{X}_1) \leq \delta$.

With this definition in hand, we derive the following characterization of the indistinguishability of X_0, X_1, which is a computational analogue of Inequality (2):

▶ **Theorem 1.5.** *If* X_0, X_1 *have* (s, ε) *pseudo-Hellinger distance* δ*, then for every* k*:*
1. $X_0^{\otimes k}, X_1^{\otimes k}$ *are* $\sqrt{2k\delta^2} + 2k\varepsilon$*-indistinguishable for circuits of size* s.
2. $X_0^{\otimes k}, X_1^{\otimes k}$ *are* $(1 - e^{-k\delta^2} - 2k\varepsilon)$ *distinguishable by circuits of size* $s' = O(sk/\varepsilon^6) + \mathrm{poly}(k/\varepsilon)$.

Thus, the number of samples needed to distinguish X_0, X_1 with constant advantage is $\Theta(1/\delta^2)$, where δ is the pseudo-Hellinger distance. As an immediate consequence, just like the traditional notion of Hellinger distance in the statistical distinguishability setting, our notion of pseudo-Hellinger distance gives an *instance-optimal* characterization of the computational distinguishability of any pairs of random variables. This is a key benefit of our work over prior works [9, 5].

To prove Part (1) of Theorem 1.5, we first observe that, by parts (1) and (2) of Definition 1.4 and a simple hybrid argument, the computational indistinguishability of $X_0^{\otimes k}$ and $X_1^{\otimes k}$ is bounded above by the computational indistinguishability of $\widetilde{X}_0^{\otimes k}$ and $\widetilde{X}_1^{\otimes k}$, plus $2k\varepsilon$. Next, the computational indistinguishability of $\widetilde{X}_0^{\otimes k}$ and $\widetilde{X}_1^{\otimes k}$ is upper-bounded by the total variation distance between $\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}$, which by Inequality (2) is bounded above by $\sqrt{2kd_H^2(\widetilde{X}_0, \widetilde{X}_1)} = \sqrt{2k\delta^2}$. This gives us the bound in Part (1) of the theorem.

To prove Part (2) of Theorem 1.5, we use Theorem 1.1 to obtain $\widetilde{X}_0, \widetilde{X}_1$ that are ε-indistinguishable from X_0, X_1 such that the computational distinguishability of $X_0^{\otimes k}, X_1^{\otimes k}$ by circuits of size s' is at least $d_{\mathrm{TV}}(\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}) - 2k\varepsilon$. In turn, Inequality (2) tells us that this is at least $1 - e^{-k \cdot d_{\mathrm{H}}^2(\widetilde{X}_0, \widetilde{X}_1)} - 2k\varepsilon$, which is then at least $1 - e^{-k\delta^2} - 2k\varepsilon$, where δ is the pseudo-Hellinger distance, $d_{\mathrm{H}}(\widetilde{X}_0, \widetilde{X}_1)$ over all $\widetilde{X}_0, \widetilde{X}_1$ that are ε-indistinguishable from X_0, X_1.

Additionally, using Part 4 of Theorem 1.1, we can prove that there exists a $\widetilde{X}_0$ such that $k = \Theta\left(1/d_{\mathrm{H}}^2(\widetilde{X}_0, X_1)\right)$ samples are both necessary and sufficient to distinguish X_0, X_1 with constant advantage. When X_1 is uniform on $\mathcal{X}$, squared Hellinger distance $d_{\mathrm{H}}^2(\widetilde{X}_0, X_1)$ can be related to the Rényi $\frac{1}{2}$-entropy of $\widetilde{X}_0$, defined as follows:

$$\mathrm{H}_{1/2}\left(\widetilde{X}_0\right) = 2\log_2\left(\sum_{i \in \mathcal{X}} \sqrt{\Pr[\widetilde{X}_0 = i]}\right).$$

Specifically, we have

$$d_H^2\left(\widetilde{X}_0, X_1\right) = 1 - 2^{-\frac{1}{2}\left(\log|\mathcal{X}| - \mathrm{H}_{1/2}\left(\widetilde{X}_0\right)\right)} = \Theta\left(\min\left\{\log|\mathcal{X}| - \mathrm{H}_{1/2}\left(\widetilde{X}_0\right), 1\right\}\right).$$

This allows us to create another suitable abstraction for the case where X_1 is uniform, beginning with the following definition of *pseudo-Rényi entropy*.

▶ **Definition 1.6.** *For random variable* X_0 *over* $\mathcal{X}$*,* $s \in \mathbb{N}$*,* $\varepsilon > 0$*, the* (s, ε) pseudo-Rényi $\frac{1}{2}$-entropy *of* X_0 *is the smallest* r *such that there exists a random variable* $\widetilde{X}_0$ *such that:*
1. X_0 *is* ε*-indistinguishable from* $\widetilde{X}_0$ *for circuits of size* s.
2. $H_{1/2}\left(\widetilde{X}_0\right) = r$.

This definition yields the following characterization of the indistinguishability of X_0 from the uniform distribution:

▶ **Theorem 1.7.** *If X_0 is a distribution over $\mathcal{X}$ that has (s, ε) pseudo-Rényi $\frac{1}{2}$-entropy $r = \log|\mathcal{X}| - \delta$ and $\mathcal{U}$ is the uniform distribution over $\mathcal{X}$, then for every k:*

1. *$X_0^{\otimes k}, \mathcal{U}^{\otimes k}$ are $O(\sqrt{k\delta}) + 2k\varepsilon$-indistinguishable for circuits of size s.*
2. *$X_0^{\otimes k}, \mathcal{U}^{\otimes k}$ are $(1 - e^{-\Omega(k \min\{\delta, 1\})} - 2k\varepsilon)$ distinguishable by circuits of size $s' = O(sk/\varepsilon^{12}) + \mathrm{poly}(k/\varepsilon)$.*

The number of samples needed to distinguish X_0 from uniform is $\Theta\left(\frac{1}{\delta}\right)$ where δ is the gap between the pseudo-Rényi $\frac{1}{2}$-entropy and its maximum possibly value (namely $\log|\mathcal{X}|$).

As mentioned, we also can deduce Geier's result [5] as stated above directly from Theorem 1.1. A formal statement and comparison can be found in the full version of the paper, but we note that Geier's quantitative bound (the specific polynomial loss in circuit complexity) is better than ours. In addition, Geier proves a version of the result for uniform distinguishers, whereas we only work with nonuniform distinguishers (i.e., boolean circuits). Both of these limitations come from the currently known theorems about *multicalibration*, which is the main technical tool we use (as discussed below), and it is interesting problem to obtain multicalibration theorems that provide tight quantitative bounds and yield uniform-complexity results in applications such as ours. The benefit of using multicalibration is that it provides a direct translation between the computational and information-theoretic setting (e.g. as captured in Theorem 1.1), which not only yields existing results as corollaries but also offers new ones, such as Theorem 1.5.

1.2 Asymptotic Complexity Formulations

In the foundations of cryptography, it is common to state computational indistinguishability results in terms of asymptotic polynomial complexity. In this section, we demonstrate the extensibility of our results to the asymptotic setting by presenting Theorem 1.5 in such a way, along with concrete definitions of this asymptotic regime. We start by formalizing indistinguishability in the asymptotic setting (i.e. with respect to ensembles of random variables):

▶ **Definition 1.8.** *Let $X = \{X_n\}_{n\in\mathbb{N}}$ and $Y = \{Y_n\}_{n\in\mathbb{N}}$ be ensembles of random variables, where each X_n, Y_n are supported on $\{0,1\}^{m(n)}$ and $m(n) = n^{O(1)}$. For $\eta : \mathbb{N} \to [0,1]$, we say that $X \equiv_\eta^{\mathrm{comp}} Y$ if for all c, there exists an n_0 such that for all $n \geq n_0$, X_n is $(\eta(n) + n^{-c})$-indistinguishable from Y_n by circuits of size n^c.*

Equivalently, we say $X \equiv_\eta^{\mathrm{comp}} Y$ if there exists $s(n) = n^{\omega(1)}$, $\varepsilon(n) = n^{-\omega(1)}$ such that for all n X_n is $(\eta(n) + \varepsilon(n))$-indistinguishable from Y_n with respect to size $s(n)$ circuits.[1]

For simplicity, we will also use $X \equiv^{\mathrm{comp}} Y$ to denote that there exists *some function $\eta : \mathbb{N} \to [0,1]$, $\eta(n) = n^{-\omega(1)}$ such that $X \equiv_\eta^{\mathrm{comp}} Y$.*

We also generalize the notion of pseudo-Hellinger distance (Definition 1.4) to the setting of ensembles of random variables:

▶ **Definition 1.9.** *For ensembles of random variables X and Y, we say that X and Y have pseudo-Hellinger distance at most δ (denoted $X \equiv_\delta^{cH} Y$) for a function $\delta : \mathbb{N} \to [0,1]$, if there exist ensembles $\widetilde{X} = \{\widetilde{X}_n\}, \widetilde{Y} = \{\widetilde{Y}_n\}$ such that $X \equiv^{\mathrm{comp}} \widetilde{X}$, $Y \equiv^{\mathrm{comp}} \widetilde{Y}$, and $\forall n$, $d_H(\widetilde{X}_n, \widetilde{Y}_n) \leq \delta(n)$.*

[1] The equivalence of these two formulations goes back to Bellare [1].

Finally, we require a notion of the sample complexity required to distinguish ensembles:

▶ **Definition 1.10.** *We say that ensembles X, Y have computational sample complexity at least $k : \mathbb{N} \to \mathbb{N}$ for $k(n) = n^{O(1)}$ if $X^k \equiv^{\text{comp}}_{1/2} Y^k$ (where $X^k = \{X_n^{k(n)}\}$)[2]. We denote this as $X \equiv^{cS}_{k} Y$.*

With these definitions, we can now state a corollary of our characterization of the indistinguishability of random variables in terms of their pseudo-Hellinger distance.

▶ **Corollary 1.11** (Asymptotic Formulation of Theorem 1.5)**.** *Let X, Y be ensembles of random variables, let $\delta : \mathbb{N} \to [0,1]$, let $k : \mathbb{N} \to \mathbb{N}$ be a function such that $k(n) = n^{O(1)}$, and let $\delta(n) = n^{-O(1)}$.*

Then:

1. *If $X \equiv^{cH}_{\delta} Y$, then $X \equiv^{cS}_{k} Y$ for some $k(n) = \Omega\left(\frac{1}{\delta(n)^2}\right)$.*
2. *If $X \equiv^{cS}_{k} Y$, then $X \equiv^{cH}_{\delta} Y$ for some $\delta(n) = O\left(\frac{1}{\sqrt{k(n)}}\right)$.*

We delegate the proof of this to the full version of the paper (in its appendix).

1.3 Technical Overview

Our work builds on the recent work of Casacuberta, Dwork, and Vadhan [2]. Driving inspiration from [4], they showed how the recent notion of *multicalibration* from the algorithmic fairness literature [10] leads to simpler and more illuminating proofs of a number of fundamental known results in complexity theory and cryptography, such as the Impagliazzo Hardcore Lemma [12], the Dense Model Theorem [14], and characterizations of pseudoentropy [16]. We show how multicalibration can be used to derive *new* results about computational indistinguishability (Theorems 1.1 and 1.5) and in general reduce computational reasoning to information-theoretic reasoning. Specifically, applying the Multicalibration Theorem [10] to distinguishing problems in a way inspired by [2], we obtain the following. Let $\widetilde{X}_{b|p(\widetilde{X}_b)=y}$ be the distribution where x is sampled according to $\widetilde{X}_b$ conditioned on $p(x) = y$, for some function $p : \mathcal{X} \to [m]$. Let $p(X_0)$ be the distribution over $[m]$ where for $y \in [m]$, the probability that $p(X_0) = y$ is exactly $\sum_{x \in p^{-1}(y)} X_0(x)$.

▶ **Theorem 1.12.** *For every pair of random variables X_0, X_1, every positive integer s, and every $\varepsilon > 0$, there exist random variables $\widetilde{X}_0, \widetilde{X}_1$ and a function $p : \mathcal{X} \to [m]$ for $m = O(1/\varepsilon)$ such that:*

- **(a)** *$\widetilde{X}_0$ is ε-indistinguishable from X_0 and $\widetilde{X}_1$ is ε-indistinguishable from X_1 for circuits of size s.*
- **(b)** *$p(X_0)$ is identically distributed to $p(\widetilde{X}_0)$ and $p(X_1)$ is identically distributed to $p(\widetilde{X}_1)$.*
- **(c)** *For every $y \in [m]$, $\widetilde{X}_{0|p(\widetilde{X}_0)=y}$ is identically distributed to $\widetilde{X}_{1|p(\widetilde{X}_1)=y}$.*
- **(d)** *p is computable by circuits of size $O(s/\varepsilon^6)$.*

Because the above theorem shows that $\widetilde{X}_0, X_0$ and $\widetilde{X}_1, X_1$ are ε-indistinguishable for circuits of size s, then by a simple hybrid argument, $\widetilde{X}_0^{\otimes k}, X_0^{\otimes k}$ and $\widetilde{X}_1^{\otimes k}, X_1^{\otimes k}$ are εk-indistinguishable for circuits of size s. Thus, one direction of Theorem 1.1 clearly follows given Theorem 1.12: size s circuits cannot distinguish $X_0^{\otimes k}, X_1^{\otimes k}$ better than $d_{\text{TV}}(\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}) + 2k\varepsilon$.

[2] This choice of 1/2 is arbitrary, and can be amplified through repetition.

However, the other direction is more subtle; namely, showing that one *can* distinguish $X_0^{\otimes k}, X_1^{\otimes k}$ with advantage approaching $d_{\mathrm{TV}}(\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k})$ with small circuits. For this, we heavily rely on items (b), (c), and (d) from Theorem 1.12, as well as the property that $m = O(1/\varepsilon)$. The intuition is the following: consider an element x that is sampled from either $\widetilde{X}_0, \widetilde{X}_1$, and let $p(x) = y$. By item (c), we know that once we condition on a value $p(x) = y$, $\widetilde{X}_0, \widetilde{X}_1$ are identically distributed. Thus, there is no information to be gained from the sample x *besides* the label of the partition that x is in (i.e., the value $p(x)$). In fact, this gives us a simple recipe for the *optimal* distinguisher between $\widetilde{X}_0, \widetilde{X}_1$: for each sample x we see, let us compute the value $p(x) = y$, and see how much more likely it is to sample an element in $p^{-1}(y)$ under $\widetilde{X}_0$ vs. $\widetilde{X}_1$. We can simply keep a running product over all of the samples $x_1, \ldots, x_k$ of the form

$$\prod_{j=1}^{k} \frac{\mathbb{P}_{x \sim \widetilde{X}_1}[p(x) = p(x_j)]}{\mathbb{P}_{x \sim \widetilde{X}_0}[p(x) = p(x_j)]}.$$

If the above product is larger than 1, this means that a sequence of values is more likely under $\widetilde{X}_1$, and if it is less than 1, then this means a sequence of partitions of more likely under $\widetilde{X}_0$. In particular, calculating this simple expression and checking whether it is larger than 1 is an *optimal* distinguisher for $\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}$, as it is just the maximum likelihood decoder (that is to say, this decoder achieves distinguishing advantage $d_{\mathrm{TV}}(\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k})$ between $\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}$). All that remains is to show that we can encode this distinguisher using small circuits. Here, we rely on point (d) of Theorem 1.12: computing the function p can be done with circuits of small size. Thus, for any element x, we compute the index of the partition that it is in $p(x)$, and feed this index into a look-up table (encoded non-uniformly) such that it returns the value $\frac{\mathbb{P}_{x \sim \widetilde{X}_1}[p(x)=p(x_j)]}{\mathbb{P}_{x \sim \widetilde{X}_0}[p(x)=p(x_j)]}$. Across all k elements we see, we then must only keep track of the product, and return whether the value it computes is ≥ 1. Here, we use the fact that p has a small domain to prove that we can actually encode the look-up table mapping $y \in [m]$ to the values $\frac{\mathbb{P}_{x \sim \widetilde{X}_1}[p(x)=y]}{\mathbb{P}_{x \sim \widetilde{X}_0}[p(x)=y]}$ without using too large of a circuit. Of course, there are also issues with overflow and underflow in performing the numerical calculations, but we show this can be computed formally with small circuits in the full version of the paper.

Finally, it remains to analyze the distinguisher we have constructed here for $\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}$ when we apply it to $X_0^{\otimes k}, X_1^{\otimes k}$. Here, we rely on Part (b) of Theorem 1.12. Indeed, the distinguisher relies *only* on the probabilities $p(\widetilde{X}_0), p(\widetilde{X}_1)$ (i.e., the probability mass placed on each value of y). *But*, the distributions X_0, X_1 place exactly the same probability mass on each part of the partition (formally, $p(X_b) = p(\widetilde{X}_b)$). Thus, whatever distinguishing advantage our distinguisher achieves on $\widetilde{X}_0^{\otimes k}, \widetilde{X}_1^{\otimes k}$ is *exactly the same* as its distinguishing advantage over $X_0^{\otimes k}, X_1^{\otimes k}$. This then yields the second part of Theorem 1.1.

1.3.1 Multicalibrated Partitions

We now review the concept of multicalibration and describe how we prove Theorem 1.12 from the Multicalibration Theorem.

Multicalibration is a notion that first arose in the algorithmic fairness literature. Roughly speaking, its goal is, given a predictor h of an unknown function g, and a domain $\mathcal{X}$ (which is often thought of as a population of individuals), to ensure that every desired *subpopulation* $S \subseteq \mathcal{X}$ receives calibrated predictions from h. The subpopulations are specified by their characteristic functions f, which came from a family $\mathcal{F}$. More formally, we say that for domain $\mathcal{X}$ and distribution $\mathcal{D}$ over $\mathcal{X}$, a function $g : \mathcal{X} \to [0, 1]$, and a class $\mathcal{F}$ of functions

$f : \mathcal{X} \to [0,1]$, h is a $(\mathcal{F}, \varepsilon)$ *multicalibrated predictor* for g with respect to $\mathcal{D}$, if $\forall f \in \mathcal{F}$, and $\forall v \in \text{image}(h)$:

$$|\mathbb{E}_{x\sim\mathcal{D}}[f(x) \cdot (g(x) - h(x))|h(x) = v]| \leq \varepsilon.$$

The sets $\{h^{-1}(v) : v \in \text{Image}(h)\}$ define a partition of $\mathcal{X}$, which gives rise to the following more convenient formulation. Let $\mathcal{D}|_P$ denote the distribution $\mathcal{D}$ conditioned on being in the subset $P \subseteq \mathcal{X}$ of the domain.

We use a partition-based characterization of multicalibration as used in previous works [8, 7, 2]:

▶ **Definition 1.13** (Multicalibration [10] as formulated in [2]). *Let $g : \mathcal{X} \to [0,1]$ be an arbitrary function, $\mathcal{D}$ be a probability distribution over $\mathcal{X}$, and for an integer $s \in \mathbb{Z}^+$, let $\mathcal{F}^{(s)}$ be the class of functions $f : \mathcal{X} \to [0,1]$ computable by size s circuits. Let $\varepsilon, \gamma > 0$ be constants. Then, we say that a partition $\mathcal{P}$ of $\mathcal{X}$ is $(\mathcal{F}^{(s)}, \varepsilon, \gamma)$-approximately multicalibrated for g on $\mathcal{D}$ if for every $f \in \mathcal{F}^{(s)}$ and every $P \in \mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x \in P] \geq \gamma$ it holds that*

$$\left|\mathbb{E}_{x\sim\mathcal{D}|_P}[f(x) \cdot (g(x) - v_P)]\right| \leq \varepsilon,$$

where we define $v_P := \mathbb{E}_{x\sim\mathcal{D}|_P}[g(x)]$.

For the class $\mathcal{F}^{(s)}$ of functions computable by size s circuits, let $\mathcal{F}^{(s,k)}$ denote the set of partitions $\mathcal{P}$ such that there exists $\hat{f} \in \mathcal{F}^{(s)}$, $\hat{f} : \mathcal{X} \to [k]$ and $\mathcal{P} = \{\hat{f}^{-1}(1), \ldots, \hat{f}^{-1}(k)\}$.

The result of Hébert-Johnson, Kim, Reingold, and Rothblum [10] can be stated as follows in the language of partitions:

▶ **Theorem 1.14** (Multicalibration Theorem [10]). *Let $\mathcal{X}$ be a finite domain, for an integer $s \in \mathbb{Z}^+$, let $\mathcal{F}^{(s)}$ be the class of functions $f : \mathcal{X} \to \{0,1\}$ computable by size s circuits, let $g : \mathcal{X} \to [0,1]$ be an arbitrary function, $\mathcal{D}$ be a probability distribution over $\mathcal{X}$, and $\varepsilon, \gamma > 0$. Then, there exists a $(\mathcal{F}_s, \varepsilon, \gamma)$-approximately multicalibrated partition $\mathcal{P}$ of $\mathcal{X}$ for g on $\mathcal{D}$ such that $\mathcal{P} \in \mathcal{F}^{(W,k)}$, where $W = O(s/(\varepsilon^2\gamma^2)) + O(1/(\varepsilon^4\gamma) \cdot \log(|\mathcal{X}|/\varepsilon))$ and $k = O(1/\varepsilon)$.*

In particular, one way to understand the above definition is that the partition $\mathcal{P}$ breaks the domain into parts, such that *within* each part, for all functions $f \in \mathcal{F}$, the function g is ε-indistinguishable from its expected value.

Recently, this perspective on multicalibration has proven to be quite fruitful, with applications to machine learning [7], graph theory [4], and complexity theory and cryptography [4, 2]. Philosophically, the work of [2] showed how to reprove (and strengthen) theorems of average-case hardness and indistinguishability using multicalibration (specifically, the Impagliazzo Hardcore Lemma [12], the Dense Model Theorem [14], and characterizations of pseudoentropy [16]). One consequence of their work was a "loose" intuition that multicalibration translates questions about computational indistinguishability into questions about statistical indistinguishability. Our goal is to show this translation more clearly and generally, and underline directly how multicalibration reduces an understanding of computational indistinguishability to an understanding of statistical indistinguishability (as captured by Theorem 1.1).

1.3.2 Invoking Multicalibration

As mentioned above, multicalibration has already found a host of applications. However, for the specific purpose of distinguishing two distributions X_0, X_1, it is not immediately

clear how to apply this framework. Of course, we can naturally view each distribution as given by its probability mass function from $\mathcal{X}$ to $[0,1]$, but in order to create multicalibrated partitions, we need a function g with respect to which the partitions should be calibrated. For this, our first observation is to use the following function g:

$$g(x) = \begin{cases} 0 & \text{with probability } \frac{\mathbb{P}[X_0=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]} \\ 1 & \text{with probability } \frac{\mathbb{P}[X_1=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]}. \end{cases}$$

Roughly speaking, the function g is motivated by looking at the distribution $\mathcal{D} = \frac{1}{2}(X_0 + X_1)$. In this distribution, the probability of seeing an element x is $(\mathbb{P}[X_0 = x] + \mathbb{P}[X_1 = x])/2$. However, we can also understand the sampling procedure as first choosing one of X_0, X_1 with probability $1/2$, and then sampling from the chosen distribution. The probability that x comes from X_0 is then $\mathbb{P}[X_0 = x]/2$, and similarly for X_1. $\mathbb{E}_{x\sim\mathcal{D}}[g(x)]$ is thus measuring the relative fraction of the time that an element x comes from X_1.

Importantly, g is now a well-defined function with respect to which we can construct a multicalibrated partition. Then we will define $\widetilde{X}_b$ for $b \in \{0,1\}$ by replacing $X_b|_P$ with $\mathcal{D}|_P$.

We define this procedure formally below:

▶ **Definition 1.15.** *For a pair of random variables X_0, X_1, positive integer s, and parameter ε, we define the distribution $\mathcal{D}$, function g, random variables $\widetilde{X}_0, \widetilde{X}_1$ as follows.*

(a) *Let the distribution $\mathcal{D}$ be as follows: To sample from $\mathcal{D}$, first pick $B \sim \{0,1\}$ uniformly at random. Output a sample $x \sim X_B$. Note that*

$$\mathbb{P}_{x\sim\mathcal{D}}[x] = \frac{1}{2}\mathbb{P}[X_0 = x] + \frac{1}{2}\mathbb{P}[X_1 = x].$$

For a subset of the domain $\mathcal{S} \subseteq \mathcal{X}$, let $\mathcal{D}_{|\mathcal{S}}$ be the conditional distribution of $\mathcal{D}$ over the set $\mathcal{S}$.

(b) *We then define the randomized function g as:*

$$g(x) = \begin{cases} 0 & \textit{with probability } \frac{\mathbb{P}[X_0=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]} \\ 1 & \textit{with probability } \frac{\mathbb{P}[X_1=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]}. \end{cases}$$

(c) *Random variables $\widetilde{X}_0, \widetilde{X}_1$: Consider the multicalibrated partition $\mathcal{P} = \{P_i\}$ guaranteed by the Multicalibration Theorem when applied to the function g, distribution $\mathcal{D}$ over domain $\mathcal{X}$, class of functions $f : \mathcal{X} \to [0,1]$ computable by size s circuits, and parameters ϵ and $\gamma = \epsilon^2$. Given the multicalibrated partition, construct the random variables $\widetilde{X}_b$ as follows: to sample according to $\widetilde{X}_b$, first choose a piece of the partition $P_i \in \mathcal{P}$, where P_i is chosen with probability $\mathbb{P}[X_b \in P_i]$. Then sample $x \sim \mathcal{D}|_{P_i}$.*

With this definition, we can provide an informal overview of the proof of Theorem 1.12. We start with item (b), as its proof is essentially definitional. Recall that this item states that $p(X_0)$ and $p(\widetilde{X}_0)$ are identically distributed (and analogously for X_1). This follows because in the construction of the random variable $\widetilde{X}_0$, we sample an element from part P_i with probability $\Pr[X_0 \in P_i]$, and thus necessarily, the probability mass placed on each part P_i is identical between the two distributions. Next, we can also see that part (c) is definitional. In the construction of $\widetilde{X}_0$ and $\widetilde{X}_1$, conditioned on being in the piece P_i, each of the marginal distributions is exactly $\mathcal{D}_{P_i}$, and thus the marginal distributions match. Next, to conclude item (d) of Theorem 1.12, recall that the partition function p is exactly returned by the Multicalibration Theorem. Here, we can take advantage of the fact that multicalibrated partitions are efficiently computable given oracle access to the underlying

set of test functions (in this case, size $\leq s$ circuits). By replacing these oracles with the size $\leq s$ circuits, this yields an overall complexity of $O(s/\varepsilon^2)$ for the circuit size required to compute p. Finally, it remains to prove item (a) of Theorem 1.12. Because the partition $\mathcal{P}$ that is returned is multicalibrated with respect to our function g, by applying Theorem 2.15 of [2] and following the approach of the DMT++ proof of [2], we can show that for every part $P_i \in \mathcal{P}$ and every circuit of size $\leq s$, the distributions $X_0|_{P_i}$ and $\widetilde{X}_0|_{P_i}$ are close to indistinguishable (and likewise for X_1). By combining this "local" indistinguishability across all parts of the partition, this then yields the overall indistinguishability between $\widetilde{X}_0, \widetilde{X}_1$.

It is worth comparing our use of the Multicalibration Theorem to the use of Impagliazzo's Hardcore Theorem [12, 11] in some past works on computational indistinguishability and pseudorandomness [9, 13]. Applied to the same function g above, the Hardcore Theorem says that if X_0, X_1 are $(s', 1-\delta)$ -indistinguishable, for $s' = O(s \cdot \mathrm{poly}(1/\varepsilon, 1/\delta))$, then there is a subset $H \subseteq X$ such that $\Pr_{x \in X_0}[x \in H] = \Pr_{x \in X_1}[x \in H] = \delta$ such that $X_0|_H$ and $X_1|_H$ are (s, ε)-indistinguishable. Replacing $X_0|_H$ and $X_1|_H$ with their average as above, we obtain $\widetilde{X}_0, \widetilde{X}_1$ that are (s, ε)-indistinguishable from X_0, X_1 respectively, and such that $d_{\mathrm{TV}}(\widetilde{X}_0, \widetilde{X}_1) \leq 1 - \delta$. Thus, the Hardcore Lemma tightly captures the single-sample computational indistinguishability of X_0 and X_1 by the single-sample information-theoretic indistinguishability of $\widetilde{X}_0$ and $\widetilde{X}_1$. But it does not seem to capture enough information to obtain an instance-optimal characterization of the multiple-sample indistinguishability, as we do. Concretely, the Hardcore Lemma requires assuming some initial average-case hardness of g (which comes from the weak indistinguishability of X_0 and X_1) and only gives us indistinguishability from an information-theoretic counterpart on a small subset $H \subseteq \mathcal{X}$, which is not guaranteed to be efficiently recognizable. In contrast, the partition given by the Multicalibration Theorem covers the entire domain $\mathcal{X}$ with efficiently recognizable sets, and this is crucial for our instance-optimal characterization.

1.4 Organization of the paper

Sections 3 and additional sections from the full version prove the different components of Theorem 1.1. In Section 3, we apply the Multicalibration Theorem to prove Theorem 1.12. Theorem 1.12 Part (a) is equivalent to Theorem 1.1 Part (1). The full version presents a proof of Theorem 1.1 Parts (2) and (3), as well as a more general version of the statement we prove, which encompasses indistinguishability against families of functions beyond size s circuits. Also in the full version are proofs of Theorem 1.1 Part (4), and a general version of the statement, again for broader families of functions, which all rely on Theorem 1.12.

The full version of the paper also shows how our results and analysis imply a result comparable to the main theorem proven in [9, 5], and how it can be used to prove Theorem 1.5 and Theorem 1.7, which characterize distinguishability in terms of pseudo-Hellinger distance and pseudo-Rényi $\frac{1}{2}$-entropy (as defined in Definition 1.4 and Definition 1.6). Lastly, the full version also presents a generalization of our results to the distinguishability of general product distributions (instead of $X_0^{\otimes k}$ versus $X_1^{\otimes k}$).

2 Preliminaries

2.1 Distinguishability

Hellinger distance can be used to characterize the ability to distinguish two distributions when we receive multiple samples:

▷ Claim 2.1. [See [15, 17]]
1. $d_H^2(X, Y) \leq d_{\mathrm{TV}}(X, Y) \leq \sqrt{2} \cdot d_H(X, Y)$.

2. $d_H^2(X^{\otimes m}, Y^{\otimes m}) = 1 - (1 - d_H^2(X,Y))^m \in [1 - e^{-m \cdot d_H^2(X,Y)}, m \cdot d_H^2(X,Y)]$.

Combining the above gives us that

$$1 - e^{-m \cdot d_H^2(X,Y)} \leq d_{\mathrm{TV}}(X^{\otimes m}, Y^{\otimes m}) \leq \sqrt{2m} \cdot d_H(X,Y). \tag{3}$$

As a corollary, this implies that $m = \Theta(1/d_H^2(X,Y))$ samples are necessary and sufficient to increase the total variation distance $d_{\mathrm{TV}}(X^{\otimes m}, Y^{\otimes m})$ up to a constant.

▶ **Definition 2.2.** *Let X be a random variable over $\mathcal{X}$. The* Rényi entropy of order $1/2$ *of X is defined as:*

$$\mathrm{H}_{1/2}(X) = 2\log_2\left(\sum_{i \in \mathcal{X}} \sqrt{\Pr[X = i]}\right).$$

It can be shown that $\mathrm{H}_{1/2}(X) \leq \log_2 |\mathcal{X}|$ with equality if and only if X is uniform on $\mathcal{X}$. More generally, the gap in this inequality exactly measures the Hellinger distance to the uniform distribution on $\mathcal{X}$.

▷ Claim 2.3. Let X be a random variable over $\mathcal{X}$ and let $\mathcal{U}$ be the uniform distribution over $\mathcal{X}$. Then

$$d_H^2(X, \mathcal{U}) = 1 - \sqrt{\frac{2^{\mathrm{H}_{1/2}(X)}}{|\mathcal{X}|}}.$$

Proof. We can rewrite $d_H^2(X, \mathcal{U})$ as follows:

$$\begin{aligned} d_H^2(X, \mathcal{U}) &= \frac{1}{2}\sum_{i \in \mathcal{X}} \left(\sqrt{\mathbb{P}[X = i]} - \frac{1}{\sqrt{|\mathcal{X}|}}\right)^2 = \frac{1}{2}\left(1 + 1 - 2\sqrt{\frac{\mathbb{P}[X = i]}{|\mathcal{X}|}}\right) \\ &= 1 - \frac{1}{\sqrt{|\mathcal{X}|}} \sum_{i \in \mathcal{X}} \sqrt{\mathbb{P}[X = i]} = 1 - \sqrt{\frac{2^{\mathrm{H}_{1/2}(X)}}{|\mathcal{X}|}}. \end{aligned}$$

◁

▶ **Definition 2.4** (Statistical indistinguishability)**.** *Random variables X and Y are called ε-statistically indistinguishable if $d_{\mathrm{TV}}(X, Y) \leq \varepsilon$.*

We define computational indistinguishability for a class of functions $\mathcal{F}$:

▶ **Definition 2.5** (Computational indistinguishability)**.** *Random variables X and Y over domain $\mathcal{X}$ are ε-indistinguishable with respect to $\mathcal{F}$ if for every $f \in \mathcal{F}$:*

$$\left|\sum_{x \in \mathcal{X}} f(x) \cdot (\mathbb{P}[X = x] - \mathbb{P}[Y = x])\right| \leq \varepsilon.$$

Random variables X and Y are ε-distinguishable with respect to $\mathcal{F}$ if for every $f \in \mathcal{F}$:

$$\left|\sum_{x \in \mathcal{X}} f(x) \cdot (\mathbb{P}[X = x] - \mathbb{P}[Y = x])\right| \geq \varepsilon.$$

▶ **Definition 2.6** (Oracle-aided circuits and partitions)**.** *For a class $\mathcal{F}$ of functions, let $\mathcal{F}_{q,t}$ be the class of functions that can be computed by an oracle-aided circuit of size t with q oracle gates, where each oracle gate is instantiated with a function from $\mathcal{F}$. Let $\mathcal{F}_{q,t,k}$ denote the set of partitions $\mathcal{P}$ such that there exists $\hat{f} \in \mathcal{F}_{q,t}$, $\hat{f} : \mathcal{X} \to [k]$ and $\mathcal{P} = \{\hat{f}^{-1}(1), \ldots, \hat{f}^{-1}(k)\}$.*

Note that when we restrict our attention to $\mathcal{F}$ being the set of functions computable by size s circuits, then $\mathcal{F}_{q,t}$ is simply the set of functions computable by size $t + q \cdot s$ circuits. The discussion in the introduction is using exactly this simplification.

2.2 Multicalibration

We now recall the definition of multicalibration [10, 7, 8, 3], but state it in its full generality (with respect to arbitrary classes of functions $\mathcal{F}$):

▶ **Definition 2.7** (Multicalibration [10] as formulated in [2]). *Let $g : \mathcal{X} \to [0,1]$ be an arbitrary function, $\mathcal{D}$ be a probability distribution over $\mathcal{X}$, and $\mathcal{F}$ be a class of functions $f \in \mathcal{F}, f : \mathcal{X} \to [0,1]$. Let $\varepsilon, \gamma > 0$ be constants. A partition $\mathcal{P}$ of $\mathcal{X}$ is $(\mathcal{F}, \varepsilon, \gamma)$-approximately multicalibrated for g on $\mathcal{D}$ if for every $f \in \mathcal{F}$ and every $P \in \mathcal{P}$ satisfying $\mathbb{P}_{x\sim\mathcal{D}}[x \in P] \geq \gamma$ it holds that*

$$\left|\mathbb{E}_{x\sim\mathcal{D}|_P}[f(x) \cdot (g(x) - v_P)]\right| \leq \varepsilon,$$

where we define $v_P := \mathbb{E}_{x\sim\mathcal{D}|_P}[g(x)]$.

As mentioned above, [10] shows how to construct multicalibrated partitions with low-complexity relative to the class of functions $\mathcal{F}$. Their main result can be stated as:

▶ **Theorem 2.8** (Multicalibration Theorem [10]). *Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a class of functions, with $f \in \mathcal{F} : f : \mathcal{X} \to [0,1]$, $g : \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\varepsilon, \gamma > 0$. Then, there exists a $(\mathcal{F}, \varepsilon, \gamma)$-approximately multicalibrated partition $\mathcal{P}$ of $\mathcal{X}$ for g on $\mathcal{D}$ such that $\mathcal{P} \in \mathcal{F}_{q,t,k}$ for $t = O(1/(\varepsilon^4\gamma) \cdot \log(|\mathcal{X}|/\varepsilon))$, $q = O(1/(\varepsilon^2\gamma^2))$, and $k = O(1/\varepsilon)$.*

3 Applying the Multicalibration Theorem to Prove Theorem 1.12

In this section, we apply the Multicalibration Theorem to prove the following theorem, which is a generalization of Theorem 1.12. In later sections, we will use this theorem to prove the different components of Theorem 1.1.

▶ **Theorem 3.1** (General version of Theorem 1.12). *For every pair of random variables X_0, X_1, every family $\mathcal{F}$ of functions $f : \mathcal{X} \to [0,1]$, and every $\varepsilon > 0$, there exist random variables $\widetilde{X}_0, \widetilde{X}_1$ and a function $p : \mathcal{X} \to [m]$ for $m = O(1/\varepsilon)$ such that:*

(a) *$\widetilde{X}_0$ is ε-indistinguishable from X_0 and $\widetilde{X}_1$ is ε-indistinguishable from X_1 for functions $f \in \mathcal{F}$.*

(b) *$p(X_0)$ is identically distributed to $p(\widetilde{X}_0)$ and $p(X_1)$ is identically distributed to $p(\widetilde{X}_1)$.*

(c) *For every $y \in [m]$, $\widetilde{X}_{0|p(\widetilde{X}_0)=y}$ is identically distributed to $\widetilde{X}_{1|p(\widetilde{X}_1)=y}$.*

(d) *p is computable by functions in $\mathcal{F}_{O(1/\varepsilon^6),O(1/(\varepsilon^6)\cdot\log(|\mathcal{X}|)\cdot\log(|\mathcal{X}|/\varepsilon))}$.*

Definitions and notation used throughout the proof

The random variables $\widetilde{X}_0, \widetilde{X}_1$ and function $p : \mathcal{X} \to [m]$ in Theorem 3.1 are constructed as follows, by drawing a connection to the multicalibrated partition guaranteed by the Multicalibration Theorem.

▶ **Definition 3.2.** *(General version of Definition 1.15) For a pair of random variables X_0, X_1, class $\mathcal{F}$ of functions, and parameter ε, we define the family of functions $\mathcal{F}'$, distribution $\mathcal{D}$, function g, random variables $\widetilde{X}_0, \widetilde{X}_1$, and function $p : \mathcal{X} \to [m]$ as follows.*

(a) *Let $\mathcal{F}'$ be a family of functions such that $\mathcal{F}_{c,c\log|\mathcal{X}|} \subseteq \mathcal{F}'$. For example, if $\mathcal{F}$ corresponds to size $\leq s$ circuits, then $\mathcal{F}'$ is the family of functions given by circuits of size at most $s \cdot c + c \cdot \log|\mathcal{X}|$, for some universal constant c.*

(b) *Let the distribution $\mathcal{D}$ be as follows: To sample from $\mathcal{D}$, first pick $B \sim \{0,1\}$ uniformly at random. Output a sample $x \sim X_B$. Note that*

$$\mathbb{P}_{x\sim\mathcal{D}}[x] = \frac{1}{2}\mathbb{P}[X_0 = x] + \frac{1}{2}\mathbb{P}[X_1 = x].$$

For a subset of the domain $\mathcal{S} \subseteq \mathcal{X}$, let $\mathcal{D}|_{\mathcal{S}}$ be the conditional distribution of $\mathcal{D}$ over the set $\mathcal{S}$.

(c) *We then define the randomized function g as:*

$$g(x) = \begin{cases} 0 & \text{with probability } \frac{\mathbb{P}[X_0=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]} \\ 1 & \text{with probability } \frac{\mathbb{P}[X_1=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]}. \end{cases}$$

(d) *Random variables $\widetilde{X}_0, \widetilde{X}_1$: Consider the multicalibrated partition $\mathcal{P} = \{P_i\}$ guaranteed by the Multicalibration Theorem when applied to the function g, distribution $\mathcal{D}$ over domain $\mathcal{X}$, class $\mathcal{F}'$ of functions $f : \mathcal{X} \to [0,1]$, and parameters ϵ and $\gamma = \epsilon^2$. Given the multicalibrated partition, construct the random variables $\widetilde{X}_b$ as follows: to sample according to $\widetilde{X}_b$, first choose a piece of the partition $P_i \in \mathcal{P}$, where P_i is chosen with probability $\mathbb{P}[X_b \in P_i]$. Then sample $x \sim \mathcal{D}|_{P_i}$.*
Let $m = |\mathcal{P}|$ be the number of parts in this partition $\mathcal{P}$.

(e) *We let $p : \mathcal{X} \to [m]$ be the function that returns which part of the multicalibrated partition $\mathcal{P}$ an element $x \in \mathcal{X}$ is in. That is, if $x \in P_i$ for $P_i \in \mathcal{P}$, $p(x) = i$.*

Given this setup of $\widetilde{X}_0$, $\widetilde{X}_1$, and p, we are now ready to prove the different parts of Theorem 3.1.

3.1 Proof of Part (a)

We first analyze the behavior of the random variables over parts of the partition given by the Multicalibration Theorem. We begin by studying the indistinguishability of $X_0|_{P_i}$ versus $X_1|_{P_i}$, relying on the following lemma from [2].

▶ **Lemma 3.3** (Lemma 2.15 [2]). *Let $\mathcal{H} = \{h : \mathcal{X} \to [0,1]\}$ be a class of functions that is closed under negation and contains the all-zero ($h(x) = 0$ for all x) and all-one ($h(x) = 1$ for all x) functions. Let $\mathcal{D}$ be a distribution over $\mathcal{X}$ and consider $\varepsilon > 0$. Let $\mathcal{H}'$ be any family of functions such that $\mathcal{H}'_{c\log|\mathcal{X}|,c} \subseteq \mathcal{H}$ for a universal constant c.*

Suppose that $g : \mathcal{X} \to [0,1]$ is identified with a randomized Boolean-valued function and is $(\mathcal{H},\varepsilon)$-indistinguishable from the constant function $v := \mathbb{E}_{x\sim\mathcal{D}}[g(x)]$. Then the distribution $\mathcal{D}_{|g(x)=1}$ is $\left(\mathcal{H}', \frac{\varepsilon}{v(1-v)}\right)$-indistinguishable from $\mathcal{D}_{|g(x)=0}$ for $\mathcal{H}'$.

▶ **Lemma 3.4.** *For random variables X_0, X_1 and family of functions $\mathcal{F}$, consider the construction of the family of functions $\mathcal{F}'$, the distribution $\mathcal{D}$, function g, and partition $\mathcal{P}$ as in Definition 3.2. For $P_i \in \mathcal{P}$, let $v_{P_i} = \mathbb{E}_{x\sim\mathcal{D}|_{P_i}}[g(x)]$.*

For all $P_i \in \mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x \in P_i] \geq \gamma$, $X_0|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{v_{P_i}(1-v_{P_i})})$-indistinguishable from $X_1|_{P_i}$.

Proof. Similarly to the approach in the proof of Theorem 5.3 (DMT++) in [2], we consider the distribution $\mathcal{D}$ that picks $B \sim \{0,1\}$ at random and outputs a sample of $x \sim X_B$, and the randomized function g that outputs B. This precisely corresponds to the choice of $\mathcal{D}, g$ corresponding to X_0, X_1 and $\mathcal{F}'$ from Definition 3.2. As in the proof of Theorem 5.3 in [2], we will show that for all $P_i \in \mathcal{P}$, $\mathcal{D}_{P_i|g(x)=0} = X_0|_{P_i}$ and $\mathcal{D}_{P_i|g(x)=1} = X_1|_{P_i}$; we can therefore

use Lemma 3.3 (setting $\mathcal{H} = \mathcal{F}'$) to transfer the indistinguishability of g from v_{P_i} over each part of the partition $P_i \in \mathcal{P}$ to the indistinguishability of $X_0|_{P_i}$ and $X_1|_{P_i}$. More specifically, the Multicalibration Theorem guarantees that $\mathcal{P}$ is a low-complexity partition with $O(1/\varepsilon)$ parts such that, on each $P_i \in \mathcal{P}$ with $\mathbb{P}_{x\sim\mathcal{D}}[x \in P_i] \geq \gamma$, g is ε-indistinguishable from the corresponding constant function $v_{P_i} := \mathbb{E}_{x\sim\mathcal{D}_{P_i}}[g(x)]$. Applying Lemma 3.3 with $\mathcal{H}$ set to be $\mathcal{F}'$ therefore implies that for each part P_i of the partition $\mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x \in P_i] \geq \gamma$, $X_1|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{v_{P_i}(1-v_{P_i})})$-indistinguishable from $X_0|_{P_i}$ for any class of functions $\mathcal{F}$ such that $\mathcal{F}_{c\log|\mathcal{X}|,c} \subseteq \mathcal{F}'$, for a universal constant c.

We work out the details for showing $\mathcal{D}_{P_i|g(x)=1} = X_1|_{P_i}$ below, and the proof of $\mathcal{D}_{P_i|g(x)=0} = X_0|_{P_i}$ follows similarly. In what follows, let $\mathbb{P}_{\mathrm{rand}(g)}$ denote that the probability of an event is taken over the randomness of the randomized Boolean-valued function g.

We see that

$$\mathbb{P}_{x\sim\mathcal{D}_{P_i|g(x)=1}}[x] = \frac{\mathbb{P}_{\mathrm{rand}(g)}[g(x)=1|x]\cdot\mathbb{P}_{x\sim D_{P_i}}[x]}{\mathbb{P}_{\mathrm{rand}(g),x\sim\mathcal{D}_{P_i}}[g(x)=1]}. \tag{4}$$

Now, the denominator equals:

$$\begin{aligned}\mathbb{P}_{\mathrm{rand}(g),x\sim\mathcal{D}|_{P_i}}[g(x)=1] &= \frac{1}{2}\frac{1}{\mathbb{P}_{\mathcal{D}}[P_i]}\sum_{x\in\mathcal{X}}\mathbb{P}_{x\sim\mathcal{D}}[x]\cdot\frac{\mathbb{P}[X_1=x]}{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]}\\ &= \frac{\mathbb{P}[X_1\in P_i]}{\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1\in P_i]}.\end{aligned}$$

Similarly, the numerator equals:

$$\mathbb{P}_{\mathrm{rand}(g)}[g(x)=1|x]\cdot\mathbb{P}_{x\sim\mathcal{D}|_{P_i}}[x] = \frac{\mathbb{P}[X_1=x]}{\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1\in P_i]}.$$

Therefore, Equation (4) gives:

$$\mathbb{P}_{x\sim\mathcal{D}_{P_i|g(x)=1}}[x] = \frac{\mathbb{P}[X_1=x]}{\mathbb{P}[X_1\in P_i]} = \mathbb{P}[X_1|_{P_i}=x],$$

which means that $\mathcal{D}_{P_i|g(x)=1}$ is equivalent to $X_1|_{P_i}$. ◀

We next prove that we can represent $\mathcal{D}_{P_i}$ as a convex combination of $X_0|_{P_i}$ and $X_1|_{P_i}$.

▶ **Lemma 3.5.** *For random variables X_0, X_1 and family of functions $\mathcal{F}$, consider the construction of the distribution $\mathcal{D}$ and partition $\mathcal{P}$ as in Definition 3.2. For every $P_i \in \mathcal{P}$, define*

$$\alpha_0(P_i) = \frac{\mathbb{P}[X_0\in P_i]}{\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1\in P_i]} \quad \textit{and} \quad \alpha_1(P_i) = \frac{\mathbb{P}[X_1\in P_i]}{\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1\in P_i]}.$$

Then $\mathcal{D}|_{P_i} = \alpha_0(P_i)X_0|_{P_i} + \alpha_1(P_i)X_1|_{P_i}$.

Proof. For $x \in P_i$,

$$\mathbb{P}_{\mathcal{D}|_{P_i}}[x] = \frac{\mathbb{P}_{\mathcal{D}}[x]}{\mathbb{P}_{\mathcal{D}}[P_i]} = \frac{\frac{1}{2}\mathbb{P}[X_0=x]+\frac{1}{2}\mathbb{P}[X_1=x]}{\frac{1}{2}\mathbb{P}[X_0\in P_i]+\frac{1}{2}\mathbb{P}[X_1\in P_i]} = \frac{\mathbb{P}[X_0=x]+\mathbb{P}[X_1=x]}{\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1\in P_i]}.$$

Relating this to the conditional distributions $X_0|_{P_i}$ and $X_1|_{P_i}$ of X_0 and X_1, this expression equals:

$$= \frac{1}{\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1\in P_i]}\cdot\left(\mathbb{P}[X_0|_{P_i}=x]\cdot\mathbb{P}[X_0\in P_i]+\mathbb{P}[X_1|_{P_i}=x]\cdot\mathbb{P}[X_1\in P_i]\right).$$

Given the definition of $\alpha_0(P_i)$ and $\alpha_1(P_i)$, we see that

$$\mathbb{P}_{\mathcal{D}|_{P_i}}[x] = \alpha_0(P_i) \cdot \mathbb{P}[X_0|_{P_i} = x] + \alpha_1(P_i) \cdot \mathbb{P}[X_1|_{P_i} = x]. \tag{5}$$

Stated more concisely, $\mathcal{D}|_{P_i} = \alpha_0(P_i)X_0|_{P_i} + \alpha_1(P_i)X_1|_{P_i}$. Note that $\alpha_0(P_i) \geq 0$, $\alpha_1(P_i) \geq 0$, and $\alpha_0(P_i) + \alpha_1(P_i) = 1$, so this is indeed a convex combination. ◀

Next we prove that X_0 is computationally indistinguishable from $\widetilde{X}_0$ and X_1 is computationally indistinguishable from $\widetilde{X}_1$ on every part of the partition $\mathcal{P}$ guaranteed by applying the Multicalibration Theorem to g. We begin by noting that $X_1|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{\alpha_1(P_i)(1-\alpha_1(P_i))})$-indistinguishable from $X_0|_{P_i}$, which is a simple consequence of the following observation.

Observation. Note that, for all $P_i \in \mathcal{P}$, $v_{P_i} = \alpha_1(P_i)$, which can be seen as follows:

$$\begin{aligned}
v_{P_i} &= \mathbb{E}_{x\sim\mathcal{D}|_{P_i}}[g(x)] = \sum_{x\in P_i} \mathbb{P}_{x\sim\mathcal{D}|_{P_i}}[x] \cdot \frac{\mathbb{P}[X_1 = x]}{\mathbb{P}[X_0 = x] + \mathbb{P}[X_1 = x]} \\
&= \sum_{x\in P_i} \frac{\mathbb{P}_{x\sim\mathcal{D}}[x]}{\mathbb{P}_{\mathcal{D}}[P_i]} \cdot \frac{\mathbb{P}[X_1 = x]}{\mathbb{P}[X_0 = x] + \mathbb{P}[X_1 = x]} \\
&= \frac{1}{2}\frac{1}{\mathbb{P}_{\mathcal{D}}[P_i]} \sum_{x\in P_i} \mathbb{P}[X_1 = x] = \frac{1}{2}\frac{1}{\frac{1}{2}\mathbb{P}[X_0 \in P_i] + \frac{1}{2}\mathbb{P}[X_1 \in P_i]}\mathbb{P}[X_1 \in P_i] \\
&= \frac{\mathbb{P}[X_1 \in P_i]}{\mathbb{P}[X_0 \in P_i] + \mathbb{P}[X_1 \in P_i]}.
\end{aligned}$$

▶ **Lemma 3.6.** *For random variables X_0, X_1 and family of functions $\mathcal{F}$, consider the construction of the family of functions $\mathcal{F}'$, the distribution $\mathcal{D}$, function g, and partition $\mathcal{P}$ as in Definition 3.2. For $P_i \in \mathcal{P}$, consider the definitions of $\alpha_0(P_i)$ and $\alpha_1(P_i)$ as in Lemma 3.5.*

For all $P_i \in \mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x \in P_i] \geq \gamma$, $X_0|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{1-\alpha_1(P_i)})$-indistinguishable from $\widetilde{X}_0|_{P_i}$ and $X_1|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{\alpha_1(P_i)})$-indistinguishable from $\widetilde{X}_1|_{P_i}$.

The idea behind the proof of this lemma is as follows. First, since for all $P_i \in \mathcal{P}$, $\widetilde{X}_0|_{P_i}$ is equivalent to $\mathcal{D}|_{P_i}$, we want to argue that $\mathcal{D}|_{P_i}$ must also be indistinguishable from $X_0|_{P_i}$ on all $P_i \in \mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x \in P_i] \geq \gamma$. This is implied by the facts that $\mathcal{D}|_{P_i}$ is a convex combination of $X_0|_{P_i}$ and $X_1|_{P_i}$ and these two random variables are indistinguishable for these $P_i \in \mathcal{P}$.

Proof. By definition of $\widetilde{X}_0$, for $x \in P_i$:

$$\mathbb{P}[\widetilde{X}_0 = x] = \mathbb{P}[X_0 \in P_i] \cdot \mathbb{P}_{\mathcal{D}|_{P_i}}[x].$$

From Lemma 3.5, this equals:

$$= \mathbb{P}[X_0 \in P_i] \cdot \left(\alpha_0(P_i) \cdot \mathbb{P}[X_0|_{P_i} = x] + \alpha_1(P_i) \cdot \mathbb{P}[X_1|_{P_i} = x]\right).$$

Define $\widetilde{X}_0|_{P_i}$ to be the random variable such that for $x \in P_i$, $\mathbb{P}[\widetilde{X}_0|_{P_i} = x] = \mathbb{P}[\widetilde{X}_0 = x | x \in P_i]$. Note that this equals $\mathbb{P}_{\mathcal{D}|_{P_i}}[x]$. Recall that, from Lemma 3.4, for all $P_i \in \mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x \in P_i] \geq \gamma$ $X_1|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{\alpha_1(P_i)(1-\alpha_1(P_i))})$-indistinguishable from $X_0|_{P_i}$.

To show that $X_0|_{P_i}$ is $(\mathcal{F}, \frac{\varepsilon}{1-\alpha_1(P_i)})$-indistinguishable from $\widetilde{X}_0|_{P_i}$ we need to bound, for every $f \in \mathcal{F}'$: $\left|\sum_{x\in P_i} f(x) \cdot \left(\mathbb{P}_{\mathcal{D}|_{P_i}}[x] - \mathbb{P}_{X_0|_{P_i}}[x]\right)\right|$.

We see that

$$\left|\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}_{\mathcal{D}|_{P_i}}[x]-\mathbb{P}_{X_0|_{P_i}}[x]\right)\right|$$
$$=\left|\sum_{x\in P_i} f(x)\cdot\left((\alpha_0(P_i)-1)\mathbb{P}_{X_0|_{P_i}}[x]+\alpha_1(P_i)\mathbb{P}_{X_1|_{P_i}}[x]\right)\right|$$
$$=\alpha_1(P_i)\left|\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}_{X_0|_{P_i}}[x]-\mathbb{P}_{X_1|_{P_i}}[x]\right)\right|\leq\frac{\varepsilon}{1-\alpha_1(P_i)}.$$

Similarly, we can also show that $X_1|_{P_i}$ is $(\mathcal{F},\frac{\varepsilon}{\alpha_1(P_i)})$-indistinguishable from $\widetilde{X}_1|_{P_i}$. ◀

We are finally ready to prove Theorem 3.1 part (a).

Proof. For random variables X_0, X_1 and family of functions $\mathcal{F}$, consider the construction of the family of functions $\mathcal{F}'$, the distribution $\mathcal{D}$, function g, and partition $\mathcal{P}$ as in Definition 3.2.

We use the analysis of indistinguishability of random variables over all $P_i\in\mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x\in P_i]\geq\gamma$ to prove the indistinguishability of X_0 and $\widetilde{X}_0$ globally over the domain. Because we have results about indistinguishability over parts of the partition that have enough weight with respect to $\mathcal{D}$, but not those whose weight is too small, we need to break up the analysis to handle both types of $P_i\in\mathcal{P}$. Let $\mathcal{P}(\gamma)$ be the set of $P_i\in\mathcal{P}$ such that $\mathbb{P}_{x\sim\mathcal{D}}[x\in P_i]\geq\gamma$.

We focus on the indistinguishability of X_0 from $\widetilde{X}_0$ in the proof. The proof of indistinguishability of X_1 from $\widetilde{X}_1$ follows by similar arguments.

$$\left|\sum_{P_i\in\mathcal{P}}\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}[X_0=x]-\mathbb{P}[\widetilde{X}_0=x]\right)\right|$$
$$\leq\sum_{P_i\in\mathcal{P}(\gamma)}\left|\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}[X_0=x]-\mathbb{P}[\widetilde{X}_0=x]\right)\right|$$

$$+\sum_{P_i\in\mathcal{P}\setminus\mathcal{P}(\gamma)}\left|\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}[X_0=x]-\mathbb{P}[\widetilde{X}_0=x]\right)\right|. \tag{6}$$

Let us focus on the first of the two summations in Equation (6). We see

$$\sum_{P_i\in\mathcal{P}(\gamma)}\left|\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}[X_0=x]-\mathbb{P}[\widetilde{X}_0=x]\right)\right|$$
$$=\sum_{P_i\in\mathcal{P}(\gamma)}\mathbb{P}[X_0\in P_i]\left|\sum_{x\in P_i} f(x)\cdot\left(\mathbb{P}[X_0|_{P_i}=x]-\mathbb{P}[\widetilde{X}_0|_{P_i}=x]\right)\right|.$$

Applying Lemma 3.6, this is:

$$\leq\sum_{P_i\in\mathcal{P}(\gamma)}\mathbb{P}[X_0\in P_i]\cdot\frac{\varepsilon}{\alpha_0(P_i)}=\sum_{P_i\in\mathcal{P}(\gamma)}\alpha_0(P_i)\cdot 2\mathbb{P}[\mathcal{D}\in P_i]\frac{\varepsilon}{\alpha_0(P_i)}=2\varepsilon.$$

Let us now focus on the second of the two summations in Equation (6). We see

$$\sum_{P_i \in \mathcal{P} \setminus \mathcal{P}(\gamma)} \left| \sum_{x \in P_i} f(x) \cdot \left(\mathbb{P}[X_0 = x] - \mathbb{P}[\widetilde{X}_0 = x] \right) \right| \leq \sum_{P_i \in \mathcal{P} \setminus \mathcal{P}(\gamma)} \left(\mathbb{P}[X_0 \in P_i] + \mathbb{P}[\widetilde{X}_0 \in P_i] \right)$$
$$= \sum_{P_i \in \mathcal{P} \setminus \mathcal{P}(\gamma)} 2\mathbb{P}[X_0 \in P_i] < 4\gamma|\mathcal{P}|,$$

where the last inequality follows from observing that $\frac{1}{2}\mathbb{P}[X_0 \in P_i] \leq \mathbb{P}[\mathcal{D} \in P_i] < \gamma$. Note that, by definition of γ and by the bounds on $|\mathcal{P}|$ from the Multicalibration Theorem, $4\gamma|\mathcal{P}|$ is $O(\varepsilon)$.

Combining the different components of the analysis above, we find that:

$$\left| \sum_{P_i \in \mathcal{P}} \sum_{x \in P_i} f(x) \cdot \left(\mathbb{P}[X_0 = x] - \mathbb{P}[\widetilde{X}_0 = x] \right) \right| \leq O(\varepsilon).$$

Therefore, X_0 is $(\mathcal{F}, O(\varepsilon))$-indistinguishable from $\widetilde{X}_0$. Similarly, X_1 is $(\mathcal{F}, O(\varepsilon))$-indistinguishable from $\widetilde{X}_1$.

Additionally, to conclude the theorem, note that by definition of $\widetilde{X}_0, \widetilde{X}_1$, for the partition $\mathcal{P}$ given by the Multicalibration Theorem that $\widetilde{X}_0, \widetilde{X}_1$ are defined according to, for all $P_i \in \mathcal{P}$, $\widetilde{X}_0|_{P_i} = \widetilde{X}_1|_{P_i}$. ◀

3.2 Proof of Parts (b, c, d)

Here, we prove the remaining properties of Theorem 3.1:

Proof. For random variables X_0, X_1 and family of functions $\mathcal{F}$, consider the construction of the family of functions $\mathcal{F}'$, the distribution $\mathcal{D}$, function g, partition $\mathcal{P}$, and function $p : \mathcal{X} \to [m]$ as in Definition 3.2. We show the following:

▷ Claim 3.7 (Part (b) of Theorem 3.1). $p(X_0)$ is identically distributed to $p(\widetilde{X}_0)$ and $p(X_1)$ is identically distributed to $p(\widetilde{X}_1)$.

Proof. We show this WLOG for $X_0, \widetilde{X}_0$. This follows by definition. Recall that to construct $\widetilde{X}_0$, we sample part P_i with probability $\Pr[X_0 \in P_i]$, and then replace the conditional distribution over P_i with $\mathcal{D}$ (as defined in Definition 3.2). Thus, for any $P_i \in \mathcal{P}$, we have $\Pr[\widetilde{X}_0 \in P_i] = \Pr[X_0 \in P_i]$. ◁

▷ Claim 3.8 (Part (c) of Theorem 3.1). For every in $y \in [m]$, $\widetilde{X}_{0|p(\widetilde{X}_0)=y}$ is identically distributed to $\widetilde{X}_{1|p(\widetilde{X}_1)=y}$.

Proof. Again, this is merely definitional. As in Definition 3.2, for any part $P_i \in \mathcal{P}$, we define

$$\widetilde{X}_{1|P_i} = \mathcal{D}_{P_i},$$

and likewise

$$\widetilde{X}_{0|P_i} = \mathcal{D}_{P_i}.$$

Thus, the two distributions have the same marginal when conditioned on being in any part P_i. We conclude by recalling that the parts P_i are exactly the pieces $P^{-1}(y)$, for $y \in [m]$, hence yielding the statement. ◁

▷ Claim 3.9 (Part (d) of Theorem 3.1). p is computable by circuits in

$$\mathcal{F}_{O(1/\varepsilon^6),O(1/(\varepsilon^6)\cdot\log(|\mathcal{X}|)\log(|\mathcal{X}|/\varepsilon))}.$$

Proof. Recall that in Definition 3.2, p is exactly the partition function that results from constructing a multicalibrated partition with parameters $\varepsilon, \gamma = \varepsilon^2$ on the family $\mathcal{F}'$ (where $\mathcal{F}' = \mathcal{F}_{c,c\log(|\mathcal{X}|)}$). As in Theorem 2.8, such a partition function can be computed by a function in

$$\mathcal{F}'_{O(1/\varepsilon^6),O(1/(\varepsilon^6)\cdot\log(|\mathcal{X}|/\varepsilon))}.$$

This yields the claim. ◁

The proof of Theorem 3.1 then follows from the union of each of the individual claims. ◀

References

1. Mihir Bellare. A note on negligible functions. *J. Cryptol.*, 15(4):271–284, 2002. doi:10.1007/S00145-002-0116-X.
2. Sílvia Casacuberta, Cynthia Dwork, and Salil Vadhan. Complexity-theoretic implications of multicalibration. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1071–1082, 2024. doi:10.1145/3618260.3649748.
3. Sílvia Casacuberta, Cynthia Dwork, and Salil P. Vadhan. Complexity-theoretic implications of multicalibration. *CoRR*, abs/2312.17223, 2023. doi:10.48550/arXiv.2312.17223.
4. Cynthia Dwork, Daniel Lee, Huijia Lin, and Pranay Tankala. From pseudorandomness to multi-group fairness and back. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3566–3614. PMLR, 2023.
5. Nathan Geier. A tight computational indistinguishability bound for product distributions. In *Theory of Cryptography Conference*, pages 333–347. Springer, 2022. doi:10.1007/978-3-031-22365-5_12.
6. Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *J. Comput. Syst. Sci.*, 28(2):270–299, 1984. doi:10.1016/0022-0000(84)90070-9.
7. Parikshit Gopalan, Adam Tauman Kalai, Omer Reingold, Vatsal Sharan, and Udi Wieder. Omnipredictors. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA*, volume 215 of *LIPIcs*, pages 79:1–79:21. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022. doi:10.4230/LIPICS.ITCS.2022.79.
8. Parikshit Gopalan, Michael P. Kim, Mihir Singhal, and Shengjia Zhao. Low-degree multicalibration. In Po-Ling Loh and Maxim Raginsky, editors, *Conference on Learning Theory, 2-5 July 2022, London, UK*, volume 178 of *Proceedings of Machine Learning Research*, pages 3193–3234. PMLR, 2022. URL: https://proceedings.mlr.press/v178/gopalan22a.html.
9. Shai Halevi and Tal Rabin. Degradation and amplification of computational hardness. In Ran Canetti, editor, *Theory of Cryptography, Fifth Theory of Cryptography Conference, TCC 2008, New York, USA, March 19-21, 2008*, volume 4948 of *Lecture Notes in Computer Science*, pages 626–643. Springer, 2008. doi:10.1007/978-3-540-78524-8_34.
10. Úrsula Hébert-Johnson, Michael P. Kim, Omer Reingold, and Guy N. Rothblum. Multicalibration: Calibration for the (computationally-identifiable) masses. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 1944–1953. PMLR, 2018. URL: http://proceedings.mlr.press/v80/hebert-johnson18a.html.

11 Thomas Holenstein. Key agreement from weak bit agreement. In Harold N. Gabow and Ronald Fagin, editors, *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 664–673. ACM, 2005. doi:10.1145/1060590.1060689.

12 Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 538–545. IEEE, 1995. doi:10.1109/SFCS.1995.492584.

13 Ueli M. Maurer and Stefano Tessaro. A hardcore lemma for computational indistinguishability: Security amplification for arbitrarily weak prgs with optimal stretch. In Daniele Micciancio, editor, *Theory of Cryptography, 7th Theory of Cryptography Conference, TCC 2010, Zurich, Switzerland, February 9-11, 2010. Proceedings*, volume 5978 of *Lecture Notes in Computer Science*, pages 237–254. Springer, 2010. doi:10.1007/978-3-642-11799-2_15.

14 Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 76–85. IEEE, 2008. doi:10.1109/FOCS.2008.38.

15 Ton Steerneman. On the total variation and hellinger distance between signed measures; an application to product measures. *Proceedings of the American Mathematical Society*, 88(4):684–688, 1983.

16 Salil Vadhan and Colin Jia Zheng. Characterizing pseudoentropy and simplifying pseudorandom generator constructions. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 817–836, 2012. doi:10.1145/2213977.2214051.

17 David Woodruff. Cs 15-859: Algorithms for big data: Lecture 9, 2019.