Finding Simple Models of Complex Objects: From Regularity Lemmas to Algorithmic Fairness

A DISSERTATION PRESENTED
BY
SÍLVIA CASACUBERTA PUIG
TO
THE DEPARTMENT OF COMPUTER SCIENCE

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
BACHELOR OF ARTS
IN THE SUBJECTS OF
MATHEMATICS AND COMPUTER SCIENCE

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

APRIL 2023

Thesis advisors: Prof. Salil P. Vadhan & Prof. Cynthia Dwork
Shadow advisor: Dr. Fabian Gundlach Sílvia Casacuberta Puig

Finding Simple Models of Complex Objects: From Regularity Lemmas to Algorithmic Fairness

Abstract

In this thesis, we study connections between the recent literature on multi-group fairness for prediction algorithms and previous well-known results in graph theory, computational complexity, additive combinatorics, information theory, and cryptography. Our starting point are the definitions of *multiaccuracy* and *multicalibration*, which have established themselves as mathematical measures of algorithmic fairness. Multicalibration guarantees accurate (calibrated) predictions for every subpopulation that can be identified within a specified class of computations, whereas multiaccuracy is a strictly weaker notion which only guarantees accuracy on average.

The task of building multiaccurate predictors is closely related to the well-known *regularity lemma*, which is an older result in computational complexity. This is a central theorem that has many important implications in different areas, including the weak Szemerédi regularity lemma in graph theory, Impagliazzo's Hardcore Lemma in complexity theory, the Dense Model Theorem in additive combinatorics, computational analogues of entropy in information theory, and weaker notions of zero-knowledge in cryptography. The relationship between multiaccuracy and the regularity lemma thus implies that a multiaccurate predictor can prove all of these fundamental theorems. By formalizing this observation, we then ask: If we start with a multicalibrated predictor instead, what strengthened and more general versions of these fundamental theorems do we obtain? Through the lenses of multi-group fairness, we are able to cast the notion of multicalibration back into the realm of complexity theory and obtain stronger and more general versions of Impagliazzo's Hardcore Lemma, characterizations of pseudoentropy, and the Dense Model Theorem. Moreover, along the way, we present a unified approach of all these fundamental theorems.

Contents

Dedicated to Raquel Coronell Uribe

A theme that cuts across many domains of computer science and mathematics is to find simple representations of complex mathematical objects such as graphs, functions, or distributions on data. These representations need to capture how the object interacts with a class of tests, and to approximately determine the outcome of these tests. For example, a scientist is trying to find a mathematical model explaining observed data about some phenomenon, such as kinds of butterflies in a forest.
A minimal criterion for success is that the model should accurately predict the results of future observations. When is this possible?

Russell Impagliazzo, 2017 talk at the Institute for Advanced Study

Prediction algorithms assign numbers to individuals that are popularly understood as individual "probabilities" —what is the probability of 5-year survival after cancer diagnosis?— and which increasingly form the basis for life-altering decisions. Drawing on an understanding of computational indistinguishability developed in complexity theory and cryptography, we introduce Outcome Indistinguishability. Ideally, the outcomes from this generative model should "look like" the outcomes produced by Nature. A predictor satisfying Outcome Indistinguishability provides a generative model that cannot be efficiently refuted on the basis of the real-life observations produced by Nature.

Dwork et al. [DKR+21]

La concepció metafísica de la bellesa condueix a un concepte fonamental del sistema: harmonia. L'harmonia, dirà Leibniz, és simplicitat en la multiplicitat, i en això mateix consisteix la bellesa (...). Així, doncs, captar la bellesa és captar les relacions que lliguen els fenòmens entre ells.

Maria Ramon Cubells [Cub17]

Acknowledgments

First and foremost, I want to thank my advisors Salil Vadhan and Cynthia Dwork for their wonderful guidance in supervising this thesis, as well as for making me grow as a researcher and as a person during my undergraduate years. Their mentorship is what has made me want to pursue a PhD in theoretical computer science, and I feel extremely grateful that I have been able to learn from their brilliance during these years. I thank Salil Vadhan for his endless patience and generosity with his time in answering all of my questions, for his inspiring and meticulous guidance, and for always helping me when I was stuck. I thank Cynthia Dwork for teaching me how we can use theoretical computer science to tackle societal problems, and for introducing me to the field of algorithmic fairness.

I am grateful to Fabian Gundlach for reading multiple drafts of this thesis and providing extremely valuable feedback. I also thank Pranay Tankala for answering my various questions, as well as others who have provided valuable insights that have helped me navigate the topic of this thesis, including Huijia Lin, Daniel Lee, and Jessica Sorrell. The UCLA Graduate Summer School on Algorithmic Fairness that I attended during Summer of 2022 shaped a lot of my thinking on the topic of multi-group fairness, and for that I thank the organizers Cynthia Dwork and Guy Rothblum, as well as the various attendees of the program. I also want to acknowledge some of the conversations that I have had with several professors during my graduate school visits about the topic of this thesis during the past month, including Omer Reingold, Aaron Roth, Jon Kleinberg, Toniann Pitassi, Aleksandra Korolova and Nika Haghtalab. I also thank Allison Choat for helping with the logistical aspects of this thesis, as well as for her joyful emails. I am also grateful to the Herchel Smith fellowship for providing funding for my thesis research during the Summer of 2022.

More broadly, I am thankful to the Harvard Theory Group for providing a stimulating and fun environment in which to learn theoretical computer science, as well as to professors Madhu Sudan, Boaz Barak, Michael Mitzenmacher, Leslie Valiant, Ariel Procaccia, and Jelani Nelson for teaching me computer science. I am also grateful to the mathematics department for teaching me so much my undergraduate years. I thank Rasmus Kyng, Seth Flaxman, Julia Hesse, and Damián Blasi for helping me find my path in computer science research and for their invaluable mentorship.

I want to thank my friends in the computer science and mathematics departments for accompanying me in this journey, including Connor Wagaman, Emin Berker, Isaac Robinson, Benji Kan, April Chen, Richard Allen, Janna Withrow, Jennifer Liang, Ted Pyne, Noah Singer, Jordan Barkin, Patrick Song, Raphael Tsiamis, Chiara Darnton, Karly Hou, and Jessica Shand. I also want to thank my friends Ilona Demler, Adolfo Roquero, Daniel Núñez, Isabel Diersen, Ilana Cohen, Ana Tió Humphrey, Aviva Ramirez, Rivka Schusterman, and Lindsey Hightower for all of their support.

I am grateful to my friends from Barcelona Míriam Rodríguez, Anna Salafranca, Cèlia Castellví, Maria Guerra, and Meritxell Vila for not forgetting about me during my years in the United States, as well as to Benedikt Schesch and Mariona Colomer. I especially want to thank Benjy Firester for all of his endless support and constant encouragement. This thesis is dedicated to Raquel Coronell Uribe, whose presence I have dearly missed during this semester and whose strength never ceases to inspire all of us.

Lastly, I want to thank my family for all of their love and for everything that they do for me. To my brother, for being my role model ever since I was little, to my father, for instilling in me a passion for mathematics, and to my mother, for constantly believing in me.

I

Multicalibration Set-up

1 Introduction

Undeniably, algorithms are informing decisions that reach ever more deeply into our lives, from news article recommendations to criminal sentencing decisions to healthcare diagnostics. This progress, however, raises (and is impeded by) a host of concerns regarding the societal impact of computation.

TOC for Fairness: a Simons Collaboration Project

In recent years, algorithms are increasingly informing decisions that can deeply affect our lives. Propelled by the rapid progress in machine learning and the collection of vast amounts of data from individuals, algorithms are being deployed in all major spheres of society to assist in the decision-making process. For example, algorithms are currently used to decide whether someone should receive a loan, get hired for a job, receive a certain prison sentence, get accepted into a university program, or receive a particular medical treatment [ONe17], among many other use cases. Given the widespread use of prediction algorithms (i.e., an algorithm that assigns "scores" to individuals), the deployment of such technology must be done responsibly and ethically. Regulators and policy-makers have started to acknowledge this need, by pushing initiatives such as the EU AI Act [LWM22], New York City's AI Bias law [WAB+19], or the recent US AI Bill of Rights [HF22].

Biases in prediction algorithms. A major concern that arises in this context is whether or not prediction algorithms are *fair* across different subpopulations and minority groups. This is a crucial evaluation metric because research has shown time and time again that algorithms can indeed demonstrate bias against certain subgroups. In 2017, Buolamwini and Gebru evaluted the bias in automated facial analysis algorithms and datasets with respect to phenotypic subgroups, and found that the accuracy of the algorithm was almost perfect on white men but very low on black women [BG18]. Amazon was reported to terminate an internal AI recruiting tool that was shown to be biased against women [Vin18], and a ProPublica study found that the COMPAS recidivism prediction system predicts higher risks of recividism for black defendants [LMKA16].

Even though biased AI systems are a reality that pose a major societal problem, addressing it from a technical standpoint is a big challenge that is far from being solved. Given a prediction

algorithm, how can we quantitatively detect whether or not it is fair? This question has motivated numerous research directions and fairness metrics.

The emerging field of algorithmic fairness. The field of algorithmic fairness, which started to develop about a decade ago, aims to mathematically define what it means for an algorithm to be *fair*. One of the first papers in the field was *Fairness Through Awareness*, which formalized the principle that *similar individuals should be treated similarly* using a Lipschitz condition on the classifier [DHP+12]. The literature has since then exploded with numerous definitions, which roughly fall into either *individual* or *group* notions of fairness. The Fairness Through Awareness work is an example of an individual fairness notion, whereas in group fairness notions we identify a particular protected group and some statistic, and we guarantee that the statistic holds *both* in the protected group and in the overall population [HKRR18].

However, both approaches have important drawbacks; for example, [DHP+12] already showed how group fairness notions are sensitive to "fairness gerrymandering", where we are able to artificially satisfy the mathematical definition. They give one such simple example: if a steak house does not want members of a minority group to come to their establishment, yet they are required by law to advertise to, say, a 20% of members of the minority group, they can choose to advertise only to members of the minority group who are vegetarian. Then, they satisfy the fairness definition to advertise to a 20%, yet continue to be discriminatory against the minority group. Moreover, Chouldechova showed that some of these notions are mathematically incompatible [Cho17]. For example, any two out of three among Demographic Parity, Equalized Odds, and Predictive Rate Parity are incompatible with each other [BHN17]. In the case of individual fairness notions, their major drawback is that they require specifying a metric to measure closeness between individuals [Ilv19].

The multi-group framework and multicalibration. The *multi-group* framework was proposed as a way to bridge the individual and group fairness notions [HKRR18; KNRW18]. The underlying principle is the following: we want to establish some fairness group notion that not only one subgroup will satisfy, but rather that *every identifiable subgroup* will satisfy. This versatile framework allows us to consider the intersection of different minority subgroups, which represents reality more accurately (e.g., the intersection of gender, race, and socioeconomic subgroups). The notion of a *multicalibrated* predictor was introduced in 2018 within this multi-group framework, and has established itself as a increasingly popular measure of algorithmic fairness [HKRR18]. A multicalibrated predictor guarantees accurate (calibrated) predictions for every identifiable subpopulation within a collection $\mathcal{C}$ of subpopulations.

The intuition behind its definition is more clearly explained with an example from the statistical forecasting literature [Daw85]: suppose we have a predictor that predicts the probability of rain on every day. On average, the predictor performs accurately. However, suppose that when we condition on the day of the week being Sunday, then the accuracy significantly drops. Or, when we condition on the day of the month being a prime number, the accuracy significantly drops. This would indicate that, even though the predictor is accurate on average, it is not *calibrated*. Notice that this is exactly the problem that was observed in the *Gender Shades* work by Buolamwini and Gebru: even though the facial recognition system performs well on average, when conditioned on individuals who are black women, the accuracy significantly drops [BG18]. The notion of multicalibration precisely prohibits this accuracy drop when conditioning the predictor on any

protected subgroup. Given that protected subgroups can be hard to identify in practice, the idea behind multicalibration is to try to ensure the calibration property for as many groups as possible, hoping that these will include many minority subgroups and their intersections. Hence, the more expressive $\mathcal{C}$ is, the better fairness guarantees we achieve.

Connections to multicalibration. The notion of multicalibration has turned out to be very useful in practice, as well as having profound mathematical connections to other notions in computer science. For example, fairness concerns are very present in randomized controlled trials in medicine, given that minority subgroups tend to be less present in these medical study, where it harder to ensure fair representation among participants in the trial. This can cause diseases to go underdiagnosed for certain groups of patients (e.g., for women and elderly patients) [Kra18]. Barda et al. used a multicalibration boosting algorithm as a post-processing technique to improve the accuracy of minority groups *after* the randomized controlled trial data was collected [BYR+21].

On the theory side, multicalibration has shown surprising connections and applications in many areas of computer science. For example, the work on *Omnipredictors* uses the framework of multicalibration to perform loss minimization in machine learning that simultaneously works for a huge family of loss functions [GKR+21], and the work on *Universal Adaptability* uses multicalibration to adapt statistical findings to a large family of target distributions [KKG+22]. Multicalibration, and the weaker counter-part notion of *multiaccuracy* [HKRR18], have also shown to be closely related to the recently-proposed notion of *Outcome Indistinguishability* [DKR+21]. Motivated by the areas of complexity theory and cryptography, predictors that are Outcome Indistinguishable yield a generative model for outcomes that cannot be efficiently refuted on the basis of the real-life observations [DKR+21]. The motivation behind the notion of Outcome Indistinguishability (OI) is the following: if we were an algorithmic decision board (e.g., the equivalent of an FDA for algorithms, where we receive algorithmic predictors and we have to determine whether they are biased or not), how much access should we have to the predictor? For example, would we (the board) require access to the *code* of the predictor, or just samples are enough for the evaluation? This could have major implications, given that private companies usually do not disclose the code (or the data) that was used to train predictors. In the OI paper, it is shown that passing certain tests by the "FDA board" corresponds to guaranteeing that the predictor is multiaccurate or multicalibrated, where the tests that correspond to multicalibration are more stringent than those that correspond to multiaccuracy (as we would expect, given that multiaccuracy is a weaker notion) [DKR+21].

1.1 Our contributions

Our work. Motivated by all these connections, in this thesis we continue to study the mathematical foundations of multicalibration and multi-group fairness, and we formalize deep connections to well-established results in many areas of theoretical computer science that a priori have no relationship to the new field of algorithmic fairness. In particular, the multiaccuracy (MA) guarantee is essentially equivalent to the so-called *regularity lemma*, which is a major result in computer science [TTV09; JP14; CCL18]. Informally, the regularity lemma states that, given an arbitrarily complex function g, we can build a "simple" function h that "looks like" g. This is formalized using the notion of indistinguishability with respect to a class of functions $\mathcal{F}$.

The regularity lemma has connections to various fundamental and well-established theorems in

different areas of theoretical computer science. Vadhan et al. [TTV09] observed that the regularity lemma implies Szemerédi's Weak Regularity Lemma [FK99], the Dense Model Theorem [RTTV08; GT08; Tao07], and Impagliazzo's Hardcore Lemma [Imp95]. Vadhan and Zheng presented characterizations of pseudoentropy using this framework [VZ12], and Chung, Lui, and Pass presented an interactive version of LSL in the context of zero-knowledge proofs in cryptography [CLP15]. Given the "equivalence" between the algorithmic fairness notion of multiaccuracy and the regularity lemma, we see that a multiaccurate predictor can prove all of these theorems. By formalizing this observation, we then ask: *If we start with a multicalibrated predictor instead, what strengthened and more general versions of these fundamental theorems do we obtain?* Then, through the lenses of multi-group fairness, we are able to obtain stronger versions of these theorems. This question was recently explored in the case of Szemerédi regularity lemma in the work of [DLLT23]; in this thesis, we focus on Impagliazzo's Hardcore Lemma, characterizations of pseudoentropy, and the Dense Model Theorem.

Our results thus use the tools that have been recently defined and developed in the field of algorithmic fairness and cast these back into the realm of complexity theory. In doing so, we obtain a new perspective on fundamental and long-standing theorems in theoretical computer science (i.e., those that are related to the regularity lemma), which allows us to obtain stronger and more general versions of these theorems. In particular, we obtain stronger and more general versions of Impagliazzo's Hardcore Lemma, characterizations of pseudoentropy, and the Dense Model Theorem. Our proofs are based on the observation that a multicalibrated predictor induces a partition of the domain such that each part of this partition possess some sort of "inherent" indistinguishability. This "inherent" indistinguishability is due to the properties of a multicalibrated predictor. Therefore, the results of this thesis demonstrate a deep connection between complexity theory and algorithmic fairness, and we show how this connection yields fruitful results and new insights. Moreover, throughout the exposition of this thesis, we present a unified approach of all the fundamental theorems that we consider that are implied by the regularity lemma of Trevisan et al. [TTV09].

Outline of the thesis. This thesis is divided into two parts. Part I (Multicalibration set-up) presents the necessary background for both the fairness literature and the regularity lemma. We begin by introducing the formal definition of multiaccuracy (MA) and multicalibration (MC) in Chapter 2, including the boosting proof for building MA and MC predictors. We also establish the facts from multicalibration that we will need in our subsequent proofs. In Chapter 3, we present the formal statement of the regularity lemma, along with its complexity considerations, known lower-bounds, and its connections to other theorems. We also exemplify the relationship between MA/MC and the regularity lemma by showing how MA corresponds to the Frieze-Kannan weak regularity lemma.

In Part II (Main Theorems), we present our main results; namely, stronger and more general theorems for Impagliazzo's Hardcore Lemma (Chapter 4), characterizations of pseudoentropy (Chapter 5), and the Dense Model Theorem (Chapter 6). These three theorems present a unified structure in both the statements and the proofs, which we explore in Chapter 7. Lastly, in Chapter 8, we provide directions for future research, focused on applications of the regularity lemma to the field of cryptography.

2

Multi-Group Fairness Definitions

We develop and study multicalbration —a new measure of algorithmic fairness that aims to mitigate concerns about discrimination that is introduced in the process of learning a predictor from data. Multicalibration guarantees accurate (calibrated) predictions for every subpopulation that can be identified within a specified class of computations.

Hébert-Johnson et al. [HKRR18]

We begin by introducing the notation that we will be using throughout the thesis. We will always work on a finite set $\mathcal{X}$, which we call the *domain*. For example, in the fairness setting, $\mathcal{X}$ is usually thought of as a set of individuals. In some applications, we will let $\mathcal{X} = \{0,1\}^n$; that is, the set of all 2^n n-bit strings. This is a natural choice in theoretical computer science applications.

In the fairness setting, as discussed in the introduction, we work with functions that map $\mathcal{X}$ to $[0,1]$, which are called *scoring functions*, given that they map each individual in $\mathcal{X}$ into a "score". For example, a hiring platform might assign a score between 0 and 1 to each of the candidates in the domain $\mathcal{X}$. However, one of the key points in this thesis is to demonstrate that this instantiation of $\mathcal{X}$ as individuals (possibly represented as n-bit strings) is just *one particular case* of a broader framework. In each of the chapters of this thesis, depending on the application, the domain $\mathcal{X}$ represents a different object. For example, in the case of Szemerédi regularity lemma (Chapter 3), $\mathcal{X}$ corresponds to the set of edges in a complete graph. In the case of pseudoentropy (Chapter 5), $\mathcal{X}$ corresponds to the set of n-bit strings. We will always have an underlying distribution $\mathcal{D}$ over $\mathcal{X}$, which determines how the elements x are sampled from $\mathcal{D}$. In most cases in this thesis, $\mathcal{D}$ will correspond to the uniform distribution over $\mathcal{X}$.

In the context of algorithmic fairness, each $x \in \mathcal{X}$ represents an individual. We use g to denote the "true" outcomes associated to each individual, where $\mathcal{Y}$ denotes the range of such outcomes; that is, $g\colon \mathcal{X} \to \mathcal{Y}$. In most of the applications of this thesis, the function g will actually be boolean; i.e., $g\colon \mathcal{X} \to \{0,1\}$. For example, in the case of Impagliazzo's Hardcore Lemma [Imp95] (Chapter 4), every $x \in \mathcal{X}$ is mapped to either 0 or 1. We use $h\colon \mathcal{X} \to [0,1]$ to denote the *predictor*, which gives a "score" to each individual in $\mathcal{X}$. The function h is also referred to as the *simulator*. One should

think of h as the function that is trying to "mimic" or "approximate" g. The key point is that g is allowed to be an arbitrarily complex function, while h should be of *low-complexity*. We will define this term precisely in this chapter, but the idea is that h is a "simple" approximation of the complex function g, where the notion of approximation is defined in terms of *indistinguishability* with respect to a family of distinguishers. That is, h should satisfy two properties: First, it should be "simple" (i.e., of low complexity) with respect to the class of distinguishers. Second, it should be a good model for the function g, where "good" is measured with respect to some class of distinguishers. In other words, h is "good" for g with respect to a class of distinguishers if none of the distinguishers in the class can tell the difference between the outputs produced by g and the outputs produced by h.

We use $\mathcal{F}$ to denote the class of distinguishers, which is an arbitrary set of functions $f\colon \mathcal{X} \to \{0,1\}$, where each f corresponds to a distinguisher. (The name of "distinguisher" is a typical term in theoretical computer science, but the f's in $\mathcal{F}$ are simply functions.) In most applications, the functions f are boolean-valued, although in some cases we consider real-valued distinguishers (i.e., $f\colon \mathcal{X} \to [0,1]$). The point of defining how good of an approximation h is with respect to g, or, more precisely, of defining *indistinguishability* between h and g in terms of the class $\mathcal{F}$, is that it allows us to instantiate the class $\mathcal{F}$ differently in each setting. Formally:

Definition 2.1 ($(\mathcal{F},\epsilon)$-indistinguishability [TTV09])**.** Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a class of functions $f\colon \mathcal{X} \to \{0,1\}$, $g\colon \mathcal{X} \to [0,1]$, $\mathcal{D}$ a distribution on $\mathcal{X}$ and $\epsilon > 0$. We say that a function $h\colon \mathcal{X} \to [0,1]$ is $(\mathcal{F},\epsilon)$-*indistinguishable* from g on $\mathcal{D}$ if, for all $f \in \mathcal{F}$,

$$\left| \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[f(x)\cdot(g(x)-h(x))] \right| \leq \epsilon.$$

We remark that $x \sim \mathcal{D}$ indicates that x is sampled from $\mathcal{X}$ according to distribution $\mathcal{D}$. Moreover, by linearity of expectation, the condition stated in Definition 2.1 is equivalent to requiring the absolute value of the difference between $\mathbb{E}_{x\sim\mathcal{D}}[f(x)g(x)]$ and $\mathbb{E}_{x\sim\mathcal{D}}[f(x)h(x)]$ to be small. Naturally, we are interested in the cases where g is a "complex" function, whereas h is a "simple" function. (Otherwise, setting $h := g$ trivially satisfies the above definition.) In the case of algorithmic fairness, g is complex because it corresponds to the "true" predictions in real-life (e.g., will someone become sick), which can be arbitrarily complex (see [DKR+21] for an extended discussion of these complexity considerations in the setting of algorithmic fairness).

The difference between a "complex" and a "simple" function is quantified in terms of *complexity*, and a recurring idea that we will come back to throughout the thesis is that of the *complexity* of $\mathcal{F}$. Intuitively, we think of the complexity of $\mathcal{F}$ as how "hard" it is to compute the functions f in $\mathcal{F}$. Normally, in theoretical computer science, the notion of *circuit complexity* is used to formally capture this idea: the complexity of a function corresponds to the size of a circuit that computes it. A recurring idea in theoretical computer science is that restricting this complexity in some ways can enable results that would otherwise be impossible. For example, in cryptography, we usually work with polynomially-sized distinguishers, instead of computationally unbounded distinguishers, and many fruitful results have arisen due to this important distinction.

As in the case of the domain $\mathcal{X}$, each application in this thesis will require instantiating $\mathcal{F}$ accordingly. For example, in the case of Szemerédi regularity lemma for graphs on n vertices (Chapter 3), the domain $\mathcal{X}$ corresponds to $[\binom{[n]}{2}]$; that is, the set of pairs of elements from $[n]$. In

turn, the set $\mathcal{F}$ corresponds to a set of indicator cut functions in a graph; i.e., for every disjoint set of vertices S, T of the graph, $\mathcal{F}$ corresponds to the set of functions $f_{S,T}$ corresponds to the characteristic function of the set of edges having one endpoint in S and one in T [TTV09]. In other settings, $\mathcal{F}$ corresponds to the family of functions computable by a circuit of size s.

Remark 2.2. Throughout this thesis, we always assume that the constant 0 and 1 functions are in the class $\mathcal{F}$, which we denote by $\mathbf{1}$ and $\mathbf{0}$, respectively. That is, $\mathbf{1}(x) = 1$ and $\mathbf{0}(x) = 0$ for all $x \in \mathcal{X}$.

Remark 2.3. In some applications, we will consider the closure $\mathcal{F}'$ of $\mathcal{F}$ under "negation". That is, $\mathcal{F}' := \{f, -f \mid f \in \mathcal{F}\}$.

In the fairness setting, the distinguishers always correspond to membership indicator functions, where membership is being evaluated against a set of protected subgroups. The idea is as follows: the motivation behind the notions of multiaccuracy and multicalibration is to ensure that members that belong to protected subgroups continue to receive accurate predictions. To do so, we specify a collection $\mathcal{C}$ of protected subgroups (which can intersect). Then, for each $S \in \mathcal{C}$, we consider the function $c_S \colon \mathcal{X} \to \{0, 1\}$, where $c_S(i) = 1$ if and only if $i \in S$. That is, c_S corresponds to the membership indicator function to the protected subgroup S. In the algorithmic fairness applications of multiaccuracy and multicalibration, we set $\mathcal{F}$ to contain the functions c_S for every $S \in \mathcal{C}$. In the fairness setting, we understand the arbitrarily-complex function g as the "true" probabilities in real life. For example, if we are studying a certain illness, then $g(x) \in \{0, 1\}$ represents the real-life outcome of individual x; i.e., $g(x) = 1$ indicates that individual x gets sick. Then returning to our explanation above to what a "good approximation" of g by h means in this context, it should now be clear why this indstinguishability framework helps ensure fairness guarantees: We are requiring the predictor h to be a good model for g *even when conditioning on membership to a certain protected subgroup*. In other words, if $S \in \mathcal{C}$ corresponds to the group of women, then the distinguisher c_S must be unable to distinguish between the outputs of g versus the outputs of h conditioned on the fact that the inputs to g and h are individuals who are women. Then, h must correctly predict the illnesses not just among all individuals in $\mathcal{X}$ but also among the group of women from $\mathcal{X}$.

The richness of this multi-group fairness model —namely, considering a *class* of protected subgroups which can intersect— is one of its key advantages over individual and group fairness notions. The above explanation also highlights the idea that we can always understand a boolean distinguisher as a membership indicator function of some set (i.e., the *characteristic function* of the set). The will be a recurring perspective throughout the thesis, as we will be interchanging between sets and their characteristic functions. (In particular, we can always regard a boolean distinguisher as a membership indicator function of some set.)

Summarizing, there are three main functions in play in most definitions and theorems throughout this thesis, all of which are defined over the domain $\mathcal{X}$:

$\boldsymbol{f}$	The distinguishers, which constitute the *class of distinguishers* $\mathcal{F}$
$\boldsymbol{g}$	The arbitrarily-complex function that we wish to approximate
$\boldsymbol{h}$	The predictor or simulator, which approximates g relative to $\mathcal{F}$

2.1 Multiaccuracy and multicalibration

We can now introduce the formal definitions of multiaccuracy and multicalibration, which capture the intuitive ideas discussed above. Whenever we write $x \sim \mu$ below an expectation $\mathbb{E}$ or a probability Pr, we mean that x is sampled according to the distribution μ. While the original notions of multiaccuracy and multicalibration in Hébert-Johnson et al. [HKRR18] were stated for boolean distinguishers given their algorithmic fairness setting, many subsequent works state the definition with real-valued distinguishers (e.g., [KGZ19]). We adopt this convention, given that the regularity lemma of Trevisan et al. [TTV09] that we will see in Chapter 3 which is analogous to multiaccuracy uses real-valued distinguishers f as well.

Definition 2.4 (Multiaccuracy [HKRR18; KGZ19; GKR+21])**.** Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a collection of functions $f\colon \mathcal{X} \to [0,1]$, $g\colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon > 0$. We say that $h\colon \mathcal{X} \to [0,1]$ is an $(\mathcal{F}, \epsilon)$-*multiaccurate* (MA) predictor for g on $\mathcal{D}$ if, for all $f \in \mathcal{F}$,

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}} [f(x) \cdot (g(x) - h(x))] \right| \leq \epsilon.$$

Remark 2.5. In many of our applications, the distribution $\mathcal{D}$ on $\mathcal{X}$ will be the uniform distribution on $\mathcal{X}$, in which case we will simply write $\mathbb{E}_{x \sim \mathcal{X}}$. If we do not specify what distribution $\mathcal{D}$ we are using when invoking the multiaccuracy (and multicalibration) definitions (Definition 2.4), it should be understood that $\mathcal{D}$ is implicitly taken to be the uniform distribution over $\mathcal{X}$.

Definition 2.4 guarantees that the predictor h appears unbiased according to a class of tests defined by $\mathcal{F}$. We remark that multiaccuracy is defined with respect to $\mathcal{F}, g$, ϵ, and $\mathcal{D}$.

The notion of accuracy considered in Definition 2.4 is weaker than what is referred to as "accuracy" in some machine learning contexts, given that Definition 2.4 considers accuracy *in expectation*. That is, it requires the predictions given by h, averaged over the set of $x \in \mathcal{X}$ such that $f(x) = 1$ for each $f \in \mathcal{F}$, to be close in expectation to the "true" values given by g, up to some slack. Moreover, since we always assume that the constant function $\mathbf{1}$ is in $\mathcal{F}$ (Remark 2.2), the definition of multiaccuracy also implies that the expected values of h and g averaged over the entire domain $\mathcal{X}$ must also be close up to some slack.

For this thesis, the key observation about the definition of multiaccuracy is that it corresponds *exactly* to the definition of $(\mathcal{F}, \epsilon)$-indistinguishability (Definition 2.1). That is, saying that a predictor h is $(\mathcal{F}, \epsilon)$-multiaccurate for g is equivalent to saying that the functions g and h are $(\mathcal{F}, \epsilon)$-indistinguishable. Importantly, as we will develop in the following chapter (Chapter 3), this notion of indistinguishability with respect to a class of functions $\mathcal{F}$ is a classical and fundamental concept in theoretical computer science first, which was first introduced in [TTV09] and later developed in [JP14; CCL18]. However, the stronger notion of multicalibration, which we are about to define, has only been recently defined in the context of algorithmic fairness, and does not have an analogue to older notions of indistinguishability. The goal of this thesis is precisely to study the implications of the new notion of multicalibration in the classic indistinguishability results, given that multicalibration is a strictly stronger notion than multiaccuracy.

Multiaccuracy can be a weak guarantee. While MA guarantees unbiased predictions, it is not enough to prevent discrimination in an algorithmic fairness setting. Consider the following example discussed in [HKRR18]: suppose that $g(x) = 1/2$ for all $x \in \mathcal{X}$, and h is such that $h(x) = 0$ for

half of the individuals in $\mathcal{X}$ and $h(x) = 1$ for the other half. Then, g and h have the same expected value, yet h has artificially created two subgroups with opposite outcomes, which can be deemed disciminatory.

To avoid this, we consider the expected value on each of the level sets of h.

Definition 2.6. Given a finite domain $\mathcal{X}$ and a predictor $h\colon \mathcal{X} \to [0,1]$ and $v \in [0,1]$, we let

$$X_v := \{x \in \mathcal{X} \mid h(x) = v\}.$$

We call each X_v for $v \in \text{range}(h)$ a *level set* of h.

Then, by considering all $v \in \text{range}(h)$, we obtain a partition of $\mathcal{X}$ given by the level sets X_v. For the notion of multicalibration, we want the predictor to be unbiased within each level set. We capture this notion by conditioning on the value of h:

Definition 2.7 (Multicalibration [HKRR18; GKR+21])**.** Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a collection of functions $f\colon \mathcal{X} \to [0,1]$, $g\colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon > 0$. We say that $h\colon \mathcal{X} \to [0,1]$ is an $(\mathcal{F}, \epsilon)$-*multicalibrated* (MC) predictor for g on $\mathcal{D}$ if for all $f \in \mathcal{F}$ and for all $v \in [0,1]$ such that $\Pr[h(x) = v] > 0$,

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

An intuitive way of thinking about multicalibration is that it ensures multiaccuracy on each of the level sets of h. That is, multiaccuracy is a *global* condition, which ensures indistinguishability between g and h on average over the domain $\mathcal{X}$. On the other hand, multicalibration is a *local* condition, which ensures indistinguishability between g and h within each level set. Importantly, multicalibration also ensures indistinguishability "globally", given that h being multicalibrated implies that h is multiaccurate as well. This is an important idea which we will come back to in our results in Part II. However, a key fact about the multicalibration condition (which will also be key in our subsequent proofs) is that, by definition of a level set, $h(x) = v$ for all $x \in X_v$. Therefore, we can write the MC condition as

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

In particular, this implies that the constant value $v \in [0,1]$ is within $\pm\epsilon$ of $\mathbb{E}_{x \sim D|_{X_v}}[g(x)]$, where $D|_{X_v}$ denotes the conditional distribution. In contrast, in the MA condition, we have no further control over the values of the simulator h. Being able to operate with this constant value v rather than any possible value in $[0,1]$ will be important in some of our results, such as those described in Chapter 4. Clearly, multicalibration is a stronger notion than multiaccuracy, which is what will allow us to obtain stronger versions of the fundamental theorems that are implied by multiaccuracy by using the notion of multicalibration instead.

Another helpful way of understanding calibration in the fairness setting is the following. Recall that we interpret the outputs of h as prediction probabilities; for example, we want $h(x) = 60\%$ to indicate that individual x is 60% likely to become sick. From this point of view, multicalibration requires that the predictions can be meaningfully interpreted as conditional probabilities. That is, given a set $S \in \mathcal{C}$, ϵ-calibration with respect to S requires that the average of the true probabilities

of the individuals receiving prediction v is ϵ-close to v. This perspective of fairness as providing meaning to the concept of "individual probabilities" is thoroughly discussed in the paper on *outcome indistinguishability* [DKR+21].

Level sets of h. Throughout this thesis, it will be useful to think of the multicalibration notion in terms of the *level sets* of the simulator h. Because $\mathcal{X}$ is a finite domain, the range of h is finite as well. Therefore, h induces a finite partition of the domain $\mathcal{X}$ into level sets X_v, where each $X_v \subset \mathcal{X}$ contains the points $x \in \mathcal{X}$ such that $h(x) = v$. We will continue to explore the relationship between a multicalibrated predictor and a partition of the domain $\mathcal{X}$ in Section 2.3, where we introduce the notion of a *multicalibrated partition*. As we will formalize, we will see that a multicalibrated *predictor* h induces a multicalibrated *partition*, by letting the partition correspond to the level sets of h. Having established this relationship, we can use the algorithms from the algorithmic fairness literature which show how to construct an MC predictor to show that we can construct an MC partition as well. Then, in the proofs of our new theorems in Part II, we will work directly with the definition of a multicalibrated partition, without going back to multicalibrated predictors.

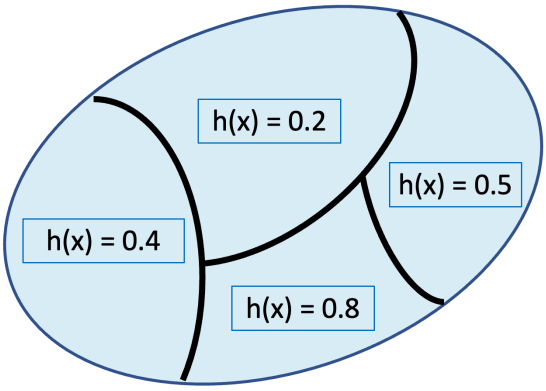

Figure 2.1: The level sets of the predictor h induce a partition of the domain $\mathcal{X}$. The definition of multicalibration asks that g and h are close in expectation within each of the level sets of h.

The size of the level sets. An important consideration regarding Definition 2.7 is that of the size of the level sets of the predictor h. While the definition of multicalibration as stated in Definition 2.7 captures the right intuition, in practice we are not able to make any guarantees about the level sets that are too small. Namely, we need to consider the quantity $\Pr_{x\sim\mathcal{D}}[x \in X_v]$ for each $v \in \text{range}(h)$. As we discuss in Section 2.4, in practice, when working in a machine learning setting, we need to learn the predictor h from samples $(x, g(x))$, and if a level set is too small then we cannot gather enough samples that for that level set, so the task becomes statistically impossible [Kim20].

There are two possible ways of relaxing the multicalibration definition in order to account for the size level sets and hence be able to build an MC predictor: One is to include a parameter γ which establishes a lower bound on the sizes of the level set that we consider. That is, the MC condition is only guaranteed for level sets that occupy at least a γ fraction of the domain $\mathcal{X}$, so that

$$\Pr_{x\sim\mathcal{D}}[x \in X_v] \geq \gamma.$$

Notice that whenever $\mathcal{D}$ corresponds to the uniform distribution (which will usually be the case in

our setting, then $\Pr_{x\sim\mathcal{D}}[x \in X_v] = |X_v|/|\mathcal{X}|$. This is why we can think of the sets that we discard as the level sets that are "too small".

This is the type of approach undertaken in [HKRR18]. Formally:

Definition 2.8 (Approximate multicalibration [HKRR18; GKR+21])**.** Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a collection of functions $f\colon \mathcal{X} \to [0,1]$, $g\colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon, \gamma > 0$. We say that $h\colon \mathcal{X} \to [0,1]$ is an $(\mathcal{F}, \epsilon, \gamma)$-*approximately multicalibrated* (MC) predictor for g on $\mathcal{D}$ if for all $f \in \mathcal{F}$ and for all $v \in [0,1]$ such that $\Pr_{x\sim\mathcal{D}}[x \in X_v] \geq \gamma$,

$$\Big| \mathop{\mathbb{E}}_{x\sim\mathcal{D}|_v} [f(x) \cdot (g(x) - h(x)) \mid h(x) = v] \Big| \leq \epsilon,$$

where $\mathcal{D}|_v$ denotes the conditional distribution $\mathcal{D}|_{h(x)=v}$ for $v \in [0,1]$ in the range of h.

Another option is to guarantee the multicalibration condition on average over the domain:

Definition 2.9 (Multicalibration on average)**.** Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a collection of functions $f\colon \mathcal{X} \to [0,1]$, $g\colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon > 0$. We say that a predictor $h\colon \mathcal{X} \to [0,1]$ is $(\mathcal{F}, \epsilon)$-*multicalibrated on average* (MCoA) for g on $\mathcal{D}$ if, for all $f \in \mathcal{F}$,

$$\mathop{\mathbb{E}}_{X_v\sim\mathcal{P}(\mathcal{D})} \Big| \mathop{\mathbb{E}}_{x\sim\mathcal{D}|_v} [f(x) \cdot (g(x) - h(x))] \Big| \leq \epsilon,$$

where $\mathcal{D}|_v$ denotes the conditional distribution $\mathcal{D}|_{h(x)=v}$ for $v \in [0,1]$ in the range of h, and $\mathcal{P}(\mathcal{D})$ denotes the distribution that selects each X_v with probability $\big(\sum_{x\in X_v} \mathcal{D}(x)\big)/\big(\sum_{x\in\mathcal{X}} \mathcal{D}(x)\big)$.

Then, when we consider a fixed level set $X_v = \{x \in \mathcal{X} \mid h(x) = v\}$, it follows that

$$\Big| \mathop{\mathbb{E}}_{x\sim\mathcal{D}|_v} [f(x) \cdot (g(x) - h(x))] \Big| \leq \frac{\epsilon}{\Pr_{\mathcal{D}}[x \in X_v]}.$$

When $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$, the multicalibration condition becomes

$$\Big| \mathop{\mathbb{E}}_{x\in X_v} [f(x) \cdot (g(x) - h(x))] \Big| \leq \epsilon \cdot \frac{|\mathcal{X}|}{|X_v|},$$

given that

$$\mathop{\mathbb{E}}_{x\in X_v} [f(x) \cdot (g(x) - h(x))] = \mathop{\mathbb{E}}_{x\sim\mathcal{X}} [f(x) \cdot (g(x) - h(x)) \mid h(x) = v]$$

and $|X_v|/|\mathcal{X}| = \Pr_{x\sim\mathcal{X}}[x \in X_v]$.

In this second approach, we allow the indistinguishability parameter to degrade with the size of the level set (the smaller X_v, the worse the guarantee becomes), but if we re-parametrize the second approach by setting $\epsilon \leftarrow \epsilon\gamma$, we obtain exactly the first definition of approximate MC, given that

$$\Big| \mathop{\mathbb{E}}_{x\in X_v} [f(x) \cdot (g(x) - h(x))] \Big| \leq \epsilon \cdot \gamma \cdot \frac{|\mathcal{X}|}{|X_v|} \leq \epsilon \cdot \frac{|X_v|}{|\mathcal{X}|} \cdot \frac{|\mathcal{X}|}{|X_v|} = \epsilon.$$

In this thesis, we will be using the first relaxation of the notion of multicalibration; namely, approximate multicalibration (Definition 2.8). However, we also include the definition of MC on average

for two reasons: first, to illustrate that there are other possible ways of relaxing the notion of multicalibration (and several others exist in the literature, such as that of "swap multicalibration" [GKR23]). Second, as we now develop, a key property that we will require of an MC predictor is that it has "not too many" level sets. In particular, for an approximation parameter ϵ, we show that there always exists an (approximated) MC predictor that only has $O(1/\epsilon)$ level sets. While this is true of both approximate MC and MCoA, we write the proof for the case of MCoA because it is shorter, and it conveys the key idea behind it (namely, this can be achieved by discretizing the domain $[0,1]$).

Low complexity of the simulator. A key property of a multicalibrated predictor that we will require throughout the thesis is that it is of *low-complexity*. We will formalize this notion in Section 2.3, where we discuss multicalibrated partitions, but there are two key intuitive ideas behind this notion. The first is that h should be of low-complexity with respect to the family of distinguishers $\mathcal{F}$. We can think of this as saying that if all of the $f \in \mathcal{F}$ are "easy" to compute, then h can only be "slightly harder" to compute that the $f \in \mathcal{F}$. In Section 2.2, we show that we can indeed build a multicalibrated predictor h that is not much harder to compute than the functions in the class of distinguishers $\mathcal{F}$. To do so, we provide a boosting algorithm to build an MC predictor h that makes oracle calls to the distinguishers.

Second, the predictor h should not have "too many" level sets, so that it induces a partition $\mathcal{P}$ on $\mathcal{X}$ such that $|\mathcal{P}|$ is not too large. In particular, we will now show that we can upper-bound the number of level sets of h by $O(1/\epsilon)$, where ϵ corresponds to multicalibration parameter. To do so, we need to introduce a discretization on the values $v \in [0,1]$. For this proof, we will make use of the second approach discussed on approximate multicalibration in this section (i.e., MCoA as in Definition 2.9), given that it makes the notation clearer.

Definition 2.10 (λ-discretization)**.** Let $0 < \lambda < 1$. The λ-*discretization* of $[0,1]$ is the set

$$\Lambda[0,1] = \left\{0, \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \ldots, \frac{n\lambda}{2}, 1\right\},$$

where n is the largest odd integer smaller than $2/\lambda$.

We will now show that given this discretization, the number of level sets of a multicalibrated predictor h can be bounded by $O(1/\lambda)$. The intuition is as follows: Given a multicalibrated predictor h with m level sets, where $m > 1/\lambda$, we round each of the values $h(x)$ to the closest point in $\Lambda[0,1]$. Then, after the rounding, h has at most $1/\lambda$ level sets, yet the rounding can only have changed each output of h by at most λ. Therefore, the multicalibration property of h is maintained after the rounding (up to an additive factor of λ). Formally:

Claim 2.11. *Let $\mathcal{X}$ be a finite domain, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to [0,1]$, let $g\colon \mathcal{X} \to [0,1]$ be an arbitrary function, and let $\epsilon, \lambda \in (0,1)$. If h is an $(\mathcal{F}, \epsilon)$-MCoA predictor for g, then there exists an $(\mathcal{F}, \epsilon + \lambda/2)$-MCoA predictor h' for g such that h' has $O(1/\lambda)$ level sets.*

Proof. We define h' as follows: for each $x \in \mathcal{X}$, the value $h'(x)$ is equal to $\text{Round}(h(x))$, where the function $\text{Round}\colon [0,1] \to \Lambda[0,1]$ maps $h(x)$ to the closest value in $\Lambda[0,1]$ (rounding up in the case of ties). By definition of λ-discretization, this implies that $|h(x) - h'(x)| \leq \lambda/2$. If we denote $X'_w = \{x \in \mathcal{X} \mid h'(x) = w\}$ for each $w \in \text{range}(h')$, then $X'_w = \cup_v X_v$, where $w - \frac{\lambda}{2} \leq v \leq w + \frac{\lambda}{2}$

with $v \in \text{range}(h)$. For each $w \in \text{range}(h')$, we have

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_w}[f(x)\cdot(g(x)-h'(x)]\right| &\le \left|\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_w}[f(x)\cdot(g(x)-h(x)]\right| + \mathop{\mathbb{E}}_{x\sim\mathcal{D}|_w}\left[f(x)\cdot|h(x)-h'(x)|\right] \\ &\le \left|\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_w}[f(x)\cdot(g(x)-h(x)]\right| + \lambda/2,\end{aligned}$$

where $\mathcal{D}|_w$ denotes $\mathcal{D}|_{h'(x)=w}$ in this case. Moreover, if we denote by $\mathcal{P}'(\mathcal{D})$ the distribution that selects X'_w for $w \in \text{range}(h')$ and we denote by $\mathcal{P}(\mathcal{D}|_w)$ the one that selects X_v for $v \in \text{range}(h)$ with $w - \frac{\lambda}{2} \le v \le w + \frac{\lambda}{2}$, then

$$\begin{aligned}\mathop{\mathbb{E}}_{X'_w\sim\mathcal{P}'(\mathcal{D})}\left|\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_w}[f(x)\cdot(g(x)-h(x)]\right| &= \mathop{\mathbb{E}}_{X'_w\sim\mathcal{P}'(\mathcal{D})}\left|\mathop{\mathbb{E}}_{X_v\sim\mathcal{P}(\mathcal{D}|_w)}\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_v}[f(x)\cdot(g(x)-h(x)]\right| \\ &\le \mathop{\mathbb{E}}_{X'_w\sim\mathcal{P}'(\mathcal{D})}\mathop{\mathbb{E}}_{X_v\sim\mathcal{P}(\mathcal{D}|_w)}\left|\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_v}[f(x)\cdot(g(x)-h(x)]\right| \\ &= \mathop{\mathbb{E}}_{X_v\in\mathcal{P}(\mathcal{D})}\left|\mathop{\mathbb{E}}_{x\sim\mathcal{D}|_v}[f(x)\cdot(g(x)-h(x)]\right| \le \epsilon.\end{aligned}$$

Therefore, h' is $(\mathcal{F}, \epsilon + \lambda/2)$-multicalibrated on average. □

In the case of MCoA, while the above argument works for any λ, for our purposes we set $\lambda := \epsilon$, which implies that if we can construct an MCoA predictor, then we can obtain an MCoA predictor with $O(1/\epsilon)$ level sets.

Remark 2.12. As stated above, in this thesis we will be using the notion of approximate MC instead of MCoA. However, using a similar discretization argument as in Claim 2.11, we can obtain an approximate MC predictor with $O(1/\epsilon)$ level sets, independent of the discretization parameter λ. Therefore, in what follows, we can always assume that a multicalibrated predictor with parameter ϵ (in the case of both approximate MC and MCoA) has $O(1/\epsilon)$ level sets.

2.2 Building multiaccurate and multicalibrated predictors

Multiaccuracy and multicalibration are both *definitions*: they are properties that a predictor h might or might not satisfy. The natural follow-up question is then is whether these two notions are efficiently realizable. That is, given a family of functions $\mathcal{F}$, an arbitrary function g, a distribution $\mathcal{D}$ and a parameter ϵ, is it feasible to build a predictor h that satisfies this definitions? As it is usually the case in theoretical computer science, by "feasible" we mean polynomial-time (we will be using *circuit size* as the complexity measure), although we will also describe the specific complexity parameters of the different algorithms to build MA and MC predictors. (As we have explained, given that multiaccuracy is a weaker notion than multicalibration, we remark that a multicalibrated predictor will also satisfy multiaccuracy, but the converse is not true [Kim20].)

Hébert-Johnson et al. were the first to answer this question in the positive in the context of algorithmic fairness [HKRR18]; since then, many algorithms for building MC predictors have recently emerged [GKSZ22; GRSW22; GHK+22; GKR23; NR23; GHK+23]. We first need to formalize this notion of *feasibility*. There are two key aspects to it: first, in order to be able to talk

about complexity, we need to formalize the model of computation. A common way of doing so in theoretical computer science is to use *circuit size* as the complexity measure. In particular, *boolean circuits* are composed of *gates* and inputs that are connected by wires. The wires carry a signal that represents either the value 0 or 1. Each gate corresponds to either the OR, AND, or NOT operation [Bar20]. For example, in the case of an OR gate, it has two incoming wires, and one or more outgoing wires. If these two incoming wires carry the signals a, b for $a, b \in \{0, 1\}$, then the signal on the outgoing wires will be $XOR(a, b)$.

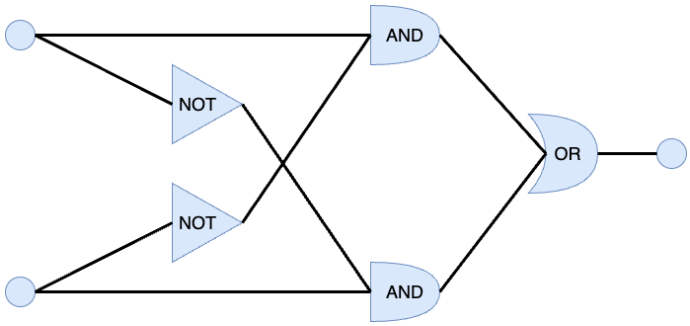

Figure 2.2: Example of a boolean circuit for computing the XOR function.

We say that a circuit C computes a function f if $f(x) = C(x)$ for each input value x. When we use the term *circuit size* for a circuit C, this corresponds to the number of gates in C.

Once we have formalized the notion of complexity, the other key notion is that of *relative* complexity. In particular, we will always consider the complexity of the predictor h *with respect to* the complexity of the class of distinguishers $\mathcal{F}$. This is formalized by allowing another type of gates in the circuits, called *oracle gates*. These are gates that are instantiated with functions f from $\mathcal{F}$. As we formalize in Definition 2.13, we count the number of oracle gates separately from the circuit size. As we will see in Algorithm 1, we will build a MA/MC by a boosting-type algorithm which repeatedly uses oracle-calls to the distinguishers $f \in \mathcal{F}$ in order to satisfy the MA/MC condition by the end.

Definition 2.13 (Relative complexity of a function [JP14, Definition 6])**.** Let $\mathcal{F}$ be a family of functions $f \colon \mathcal{X} \to [0, 1]$. A function h has *complexity* (t, q) *relative to* $\mathcal{F}$ if it 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 $\mathcal{F}$.

Definition 2.14. Given an arbitrary class of functions $\mathcal{F}$, we denote by $\mathcal{F}_{t,q}$ the class of functions that have complexity at most (t, q) relative to $\mathcal{F}$.

Remark 2.15. From Definitions 2.13 and 2.14 it follows that when working with the class $\mathcal{F}_{t,q}$ we can always assume that $q \leq t$. When we present our theorems in Part II, we will only be concerned with the parameter t (i.e., the total number of gates), in which case we will drop the second parameter q and write $\mathcal{F}_t$ as a short-hand for $\mathcal{F}_{t,\cdot}$.

Remark 2.16. In some applications, $\mathcal{F}$ is a class of distinguishers implemented by size-s circuits. In that case, every function in $\mathcal{F}_{t,q}$ can be computed by a circuit of size $t + sq$.

The notion of relative complexity captures the idea that we can make oracle calls to the functions in $\mathcal{F}$, which do not factor into the complexity. This become clearer after we analyze Algorithm 1

below. We will also then provide more intuition behind the notion of relative complexity by explaining how we can express the simulator h as a linear combination of the distinguishers $f \in \mathcal{F}$, and then the number of such distinguishers in the linear combination corresponds precisely to the number of oracle gates in the oracle-aided circuit for h.

Having formally defined the notion of relative complexity, we can now show that the notions of multiaccuracy and multicalibration are indeed realizable by a predictor h of low complexity with respect to $\mathcal{F}$.

Theorem 2.17 (Building MA predictors [HKRR18; DKR+21])**.** *For any family $\mathcal{F}$ of functions $f\colon \mathcal{X} \to [0,1]$, a function $g\colon \mathcal{X} \to [0,1]$, a distribution $\mathcal{D}$ on the domain $\mathcal{X}$, and a parameter ϵ, there exists an $(\mathcal{F}, \epsilon)$-multiaccurate predictor h such that $h \in \mathcal{F}_{t,q}$, where $t = O(1/\epsilon^2 \cdot \log(|\mathcal{X}|/\epsilon))$ and $q = O(1/\epsilon^2)$.*

Corollary 2.18. *If $\mathcal{F}$ is a class of distinguishers implemented by size-s circuits, a multiaccurate simulator h can be implemented by a circuit of size $O(s/\epsilon^2 \cdot \log(|\mathcal{X}|/\epsilon))$.*

As we discuss in the next chapter, this result turns out to be a re-discovery of the regularity lemma first shown in [TTV09], and further explored in [JP14] and [CCL18]. In both cases, we can prove the relative complexity of h via a boosting-type algorithm, as shown in Algorithm 1. As we will see throughout the thesis, this type of proof (namely, akin to boosting or gradient descent, followed by an energy potential decrease argument in the proof) can be used to prove many of the theorems that we discuss, including, for example, Impagliazzo's Hardcore Lemma and the Dense Model Theorem.

In turn, the complexity for building (approximate) multicalibrated predictors is as follows:

Theorem 2.19 (Building MC predictors [HKRR18; DKR+21])**.** *For any family $\mathcal{F}$ of functions $f\colon \mathcal{X} \to \{0,1\}$, an arbitrary function $g\colon \mathcal{X} \to [0,1]$, a distribution $\mathcal{D}$ on the domain $\mathcal{X}$, and a parameter ϵ, there exists an $(\mathcal{F}, \epsilon, \gamma)$-multicalibrated predictor h with $O(1/\epsilon)$ level sets such that $\mathcal{F}_{t,q}$, where $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$ and $q = O(1/\epsilon^2)$.*

Corollary 2.20. *If $\mathcal{F}$ is a class of distinguishers implemented by size-s circuits, then an $(\mathcal{F}, \epsilon, \gamma)$-multicalibrated simulator h can be implemented by a circuit of size $O(s/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$.*

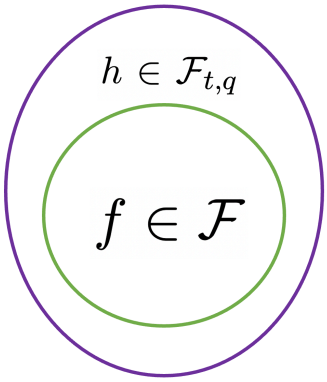

Figure 2.3: Relationship between the complexity of the distinguishers and the complexity of the predictor.

The high-level idea of how the algorithm to construct MA and MC predictors is an iterative algorithm that can be viewed as a variant of gradient descent or boosting and which works as follows [TTV09; JP14; HKRR18; DKR+21]: we start with a trivial predictor $h_0 \colon \mathcal{X} \to [0,1]$, e.g., the predictor that maps all $x \in \mathcal{X}$ to 0. Next we begin an iterative procedure where at each step t we find a distinguisher in $\mathcal{F}$ that "distinguishes" (i.e., for which the MA/MC condition does not hold) and use that distinguisher to update the predictor. The algorithm halts when there is no distinguisher that "witnesses" a violation of the MA/MC condition. Then, by definition, we have built a MA/MC predictor. The task is then to show that the algorithm does indeed terminate, and we do this by a energy potential argument. That is, we define an energy function as a function of h_t and h_{t+1} and show that the energy function strictly decreases in each iteration of the algorithm.

More concretely, the complexity stated in Theorems 2.17 and 2.19 can be obtained by roughly the following boosting-type algorithm. Recall that $\mathcal{F}'$ corresponds to the closure of $\mathcal{F}$ under "negation" (Remark 2.3).

Algorithm 1 Boosting algorithm for building a MA predictor

procedure BUILDPREDICTOR($\mathcal{F}, g, \mathcal{D}, \epsilon$) ▷ Theorems 2.17, 2.19
  Initialize $h_0 := 0$ and $t = 0$.
  while $\exists f_{t+1} \in \mathcal{F}'$ such that $\mathbb{E}_{x \sim \mathcal{D}}[f_{t+1}(x) \cdot (g(x) - h_t(x))] > \epsilon$ **do**
    $h_{t+1} := h_t + \epsilon f_{t+1}$ ▷ Update h if some f distinguishes
    $t := t + 1$
  end while
  return h_t
end procedure

In order to achieve the claimed relative complexity of h of $O(1/\epsilon^2)$ in Theorem 2.17, our goal is to show that Algorithm 1 terminates within $1/\epsilon^2$ steps. To do so, we proceed with an energy-decrease argument. We remark that this is a recurring type of proof that can be used to show most of the theorems that we consider in this thesis, as discussed in later chapters. These type of arguments proceed by defining a so-called "energy function" and then studying how this function evolves at each time step t. This is a type of proof that has been used to prove many of the theorems that we consider in this thesis.

Remark 2.21. We remark that Algorithm 1 does not ensure that the simulator h remains bounded within $[0,1]$, which is required in the multiaccuracy definition (Definition 2.4). This is why Lemma 2.22 below states the relative complexity for an MA predictor $h \colon \mathcal{X} \to \mathbb{R}$; note that we can extend the definition of multiaccuracy from $[0,1]$-valued functions to real-valued functions. However, this issue can be circumvented without increasing the relative complexity of the predictor, hence satisfying the conditions stated in Theorems 2.17, 2.19. This is shown in [TTV09, §3] and in [DKR+21]. In the latter, this is resolved by projecting onto $[0,1]$ in each step of the boosting algorithm. However, the goal of this section is to provide intuition for the relative complexity of a MA/MC, which Algorithm 1 demonstrates. We defer to the original publications for the technical details and full analysis.

Given Remark 2.21, we now prove the following:

Lemma 2.22. *Given a family $\mathcal{F}$ of functions $f\colon \mathcal{X} \to \{0,1\}$, an arbitrary function $g\colon \mathcal{X} \to [0,1]$, a distribution $\mathcal{D}$ on the domain $\mathcal{X}$, and a parameter ϵ, Algorithm 1 returns an $(\mathcal{F}', \epsilon)$-multiaccurate predictor $h\colon \mathcal{X} \to \mathbb{R}$ that can be computed with an oracle-aided circuit with $O(1/\epsilon^2)$ oracle gates, where each oracle gate is instantiated with a function from $\mathcal{F}$.*

Proof. Given the notation in Algorithm 1, we define the following energy function:

$$\Phi_t = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[(g(x) - h_t(x))^2].$$

Since h_0 is initialized at 0, the initial energy value is

$$\Phi_0 = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[(g(x) - 0)^2] = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[g(x)^2].$$

Since $g\colon \mathcal{X} \to [0,1]$ by assumption, it follows that $\mathbb{E}_{x\sim\mu}[g(x)^2] \leq 1$, and hence $\Phi_0 \leq 1$.

Next, we analyze the decrease in the potential function when we go from step t to step $t+1$:

$$\Phi_t - \Phi_{t+1} = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[(g(x) - h_t(x))^2 - (g(x) - h_t(x) - \epsilon f_{t+1}(x))^2],$$

since by the update step in Algorithm 1, $h_{t+1} = h_t + \epsilon f_{t+1}$. Then,

$$\Phi_t - \Phi_{t+1} = \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[2\epsilon \cdot f_{t+1}(x)(g(x) - h_t(x))] - \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[\epsilon^2 f_{t+1}(x)^2].$$

Lastly, since by assumption the distinguishing advantage is at least ϵ (given that the algorithm has not yet terminated), and since $f\colon \mathcal{X} \to [0,1]$, it follows that

$$\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[2\epsilon \cdot f_{t+1}(x)(g(x) - h_t(x))] - \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[\epsilon^2 f_{t+1}(x)^2] \geq 2\epsilon^2 - \epsilon^2 = \epsilon^2.$$

Therefore, we have shown that $\Phi_0 \leq 1$, $\Phi_t \geq 0$ for all t (since Φ_t is the expectation of a squared value), and $\Phi_t - \Phi_{t+1} \geq \epsilon^2$, Algorithm 1 must terminate after $O(1/\epsilon^2)$, as required. □

Returning to our discussion of the notion of *complexity of h relative to $\mathcal{F}$*, Algorithm 1 provides a very intuitive explanation for the reason why we consider oracle access to $\mathcal{F}$. The key point is that line 3 in Algorithm 1 assumes that we can find a distinguisher f for which the MA/MC condition is violated without paying any cost. In other words, this is the step where we assume oracle access to the functions in $\mathcal{F}$. Therefore, neither the size of $\mathcal{F}$ nor the complexity of the functions $f \in \mathcal{F}$ plays a role in the complexity of h relative to $\mathcal{F}$. This does play a role in other notions of complexity, as we briefly discuss in Section 2.4.

An important idea that follows from Algorithm 1 is the fact that the simulator h can be expressed as a linear combination of the distinguishers $f \in \mathcal{F}$. That is:

Remark 2.23 ([TTV09, §4])**.** The simulator h returned by Algorithm 1 can be written as a sum $h = \epsilon f_1 + \cdots + \epsilon f_q$, where $f_i \in \mathcal{F}'$. Hence, we can write

$$h = \sum_i c_i f_i, \quad \text{where} \quad \sum_i c_i^2 = q\epsilon^2.$$

This remark gives us another way of understanding why h is of low-complexity with respect to the class $\mathcal{F}$ in a way that not require using circuit complexity: namely, we can write h as a linear combination of the distinguishers $f \in \mathcal{F}$. Therefore, h is essentially as easy to compute as the distinguishers f. Indeed, by the energy-decrease argument above, we know that $k \leq \epsilon^{-2}$ in Remark 2.23. This is why this parameter q corresponds to the number of oracle gates in the circuit that computes h. (Notice that Remark 2.23 applies to the case where h is unbounded.)

In the light of Algorithm 1, we now provide a proof sketch for the parameters stated in Theorems 2.17 and 2.19.

Proof sketch for Theorem 2.17. The proof of Lemma 2.22 bounds the number of oracle calls made to the distinguishers in $\mathcal{F}$, and hence proves that $q = O(1/\epsilon^2)$ in Theorem 2.17. That is, the number of oracle gates in the oracle-aided circuit that computes h (where this circuit is built using the distinguishers determined in Algorithm 1), is equal to $O(1/\epsilon^2)$. For parameter t, namely the total number of gates in the circuit that computes h, we need to account for the rest of arithmetic operations that are required for computing h. In each of the $O(1/\epsilon^2)$ iterations in Algorithm 1, we require a scalar multiplication, a finite-precision addition, and projection onto $[0, 1]$, which can be handled by $O(\log(1/\epsilon) + \log(1/|\mathcal{X}|))$ gates. The $\log(1/\epsilon)$ term corresponds to the fact that we perform computations up to a fixed precision of $\Theta(\epsilon)$, and the $\log(1/|\mathcal{X}|)$ term corresponds to bitlength of the elements in $\mathcal{X}$. Namely, each $x \in \mathcal{X}$ requires $\log(1/|\mathcal{X}|)$ bits to represent. Hence, since there are $O(1/\epsilon^2)$ iterations in Algorithm 1, the total number of gates in the circuit computing h is $t = O(1/\epsilon^2 \cdot \log(|\mathcal{X}|/\epsilon))$, as stated in Theorem 2.17. For full details on the circuit implementation, see [DKR+21, §5]. □

Proof sketch for Theorem 2.19. The same iterative procedure in Algorithm 1 can also be used to build an approximate multicalibrated predictor (Theorem 2.19), rather than a multiaccurate one. To do so, it is enough to modify Step 3 in Algorithm 1 in order to check whether some distintinguisher distinguishes within each *level set* of the simulator h_t at step t, where we ignore the level sets that are too small. That is, Step 3 in Algorithm 1 becomes

while $\exists f_{t+1} \in \mathcal{F}'$ such that $\mathbb{E}_{x\sim\mathcal{D}}[f_{t+1}(x) \cdot (g(x) - h_t(x))) \mid h_t(x) = v] > \epsilon$ for any $v \in \Lambda[0, 1]$.

Then, the update step in Step 4 would only apply to the values of x inside the level set $X_v = \{x \in \mathcal{X} \mid h_t(x) = v\}$, while $h_{t+1}(x) = h_t(x)$ for all $x \in \mathcal{X} \setminus X_v$. A formal statement of this algorithm can be found in [HKRR18, Alg. 3.2], where they show that the total number of iterations in the energy-decrease argument becomes $O(1/(\epsilon^4\gamma))$, which corresponds to the value of parameter q in Theorem 2.19. This overhead is due to technical reasons that arise from modifying Step 3 in Algorithm 1. For parameter t, that is, the total number of gates in the circuit that computes h, the same analysis as in the proof sketch of Theorem 2.17 applies, and hence the total number of gates in the circuit computing h is $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$. For full details on the circuit implementation, see [DKR+21, §5]. □

2.3 Multicalibrated partitions

While the notion of multicalibration was originally proposed for predictors, throughout this thesis we will we will consider the notion of *multicalibrated partitions*, for reasons that will become apparent in Part II. The key idea is that the level sets of a predictor induce a partition of the domain.

Then, the level sets of a multicalibrated predictor will give us a multicalibrated partition of the domain. A similar idea was used in the work on omnipredictors [GKR+21], where the definition of a multicalibrated partition is given in terms of the covariance on each piece of the partition.Using the covariance is motivated by the literature on boosting-by-branching programs in learning theory [Kal04; KK09; MM02; KM96]. The notion of an MC partition has also been considered in [GRSW22].

Remark 2.24. Whenever we use the term "partition" $\mathcal{P}$ of the domain $\mathcal{X}$, we always imply that all the $P \in \mathcal{P}$ are pairwise disjoint and that their union is the entire domain $\mathcal{X}$.

The following formalizes the notion of a multicalibrated partition. As in the case of an MC predictor, we need to relax the definition of multicalibration for practical reasons; we choose to use the approximate MC formulation rather than the MC on average formulation to do so.

Definition 2.25 (Approximate MC partition)**.** Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a class of functions $f\colon \mathcal{X} \to [0,1]$, $g\colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon > 0$. We say that a partition $\mathcal{P}$ of $\mathcal{X}$ is $(\mathcal{F}, \epsilon, \gamma)$-*approximately multicalibrated* (MC) for g on $\mathcal{D}$ if for all $f \in \mathcal{F}$ and all $P \in \mathcal{P}$ such that $\Pr_{x\sim\mathcal{D}}[x \in P] \geq \gamma$,

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

where $v_p := \mathbb{E}_{x\sim\mathcal{D}|_P} P[g(x)]$ and $\mathcal{D}|_P$ denotes the conditional distribution $\mathcal{D}|_{h(x)\in P}$.

In the case where $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$, then

$$\Pr_{x\sim\mathcal{D}}[x \in P] = \frac{|P|}{|\mathcal{X}|}.$$

That is, in this case, Definition 2.25 should be understood as saying that we do not make any guarantees about sets that are too small (namely, about sets that occupy less than a γ fraction of the space).

In order to keep track of the impact of the size of each $P \in \mathcal{P}$, we introduce the following notation:

Definition 2.26. Given a partition $\mathcal{P}$ of the domain $\mathcal{X}$, we let $\eta_p = \Pr_{\mathcal{D}}[x \in P]$ denote the *size* parameter of $P \in \mathcal{P}$ in $\mathcal{X}$. If $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$, then $\eta_p := |P|/|\mathcal{X}|$.

That is, if P is too small (i.e., if $\eta_p < \gamma$), then Definition 2.25 makes no guarantees about this set P. This is coherent with our explanation on the size of the level sets in Section 2.1, where we discussed how the multicalibration notion cannot be satisfied in level sets that are too small.

In the theorems presented in Part II of this thesis, we will see that we will also need to establish a lower bound on the "imbalancedness" of g on each set $P \in \mathcal{P}$. Namely, we will also need to "throw away" the sets $P \in \mathcal{P}$ on which g is too "imbalanced"; i.e., such that the expected value of g on the set is too close to 0 or too close to 1. We will introduce this parameter formally in Chapter 4, which we denote by k_p.

From predictors to partitions. Next, we show how we can build a multicalibrated partition (satisfying Definition 2.25) from a multicalibrated predictor. The key intuition is the following:

Every function h induces a partition of the domain $\mathcal{X}$ given by the level sets of h. Namely, $\{X_v\}$ corresponds to a partition of $\mathcal{X}$, where $X_v = \{x \in \mathcal{X} \mid h(x) = v\}$.

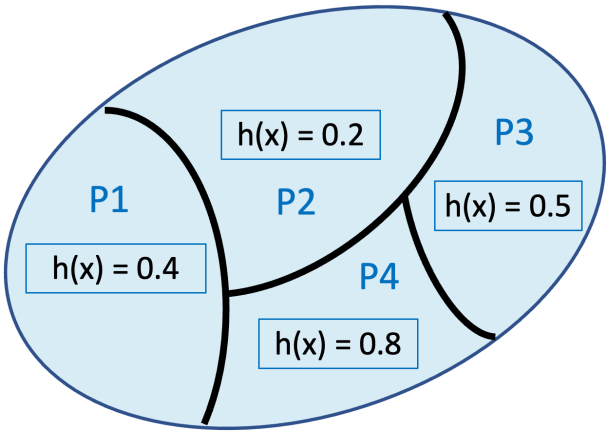

Figure 2.4: The level sets of a multicalibrated predictor induce a multicalibrated partition of the domain $\mathcal{X}$. In this case, the partition corresponds to $\mathcal{P} = \{P_1, P_2, P_3, P_4\}$ as given by the diagram.

Given this intuition, we now formalize the relationship between a multicalibrated predictor and a multicalibrated partition, using approximate multicalibration as the appropriate relaxation:

Claim 2.27. *Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a collection of functions $f\colon \mathcal{X} \to [0,1]$, $g\colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon, \gamma > 0$. If $h\colon \mathcal{X} \to [0,1]$ is an $(\mathcal{F}, \epsilon/2, \gamma)$-approximately multicalibrated predictor for g on $\mathcal{D}$, then the partition $\mathcal{P} := \{X_v\}$, where $X_v = \{x \in \mathcal{X} \mid h(x) = v\}$, is an $(\mathcal{F}, \epsilon, \gamma)$-approximately multicalibrated partition of $\mathcal{X}$ for g on $\mathcal{D}$.*

Proof. Since $h(x) = v$ for $x \in X_v$, the assumption that h is an $(\mathcal{F}, \epsilon/2, \gamma)$-MC predictor implies that

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_v} [f(x) \cdot (g(x) - v)] \right| \leq \frac{\epsilon}{2}$$

for $f \in \mathcal{F}$ and $v \in \Lambda[0,1]$ for all X_v such that $\Pr_{x\sim\mathcal{D}}[x \in X_v] \geq \gamma$. We now show that we can interchange the value v for $v_p := \mathbb{E}_{X_v}[g(x)]$. Intuitively, since multicalibration requires the expected values of g and h to be close in expectation, setting the value of h on each level set $P = X_v$ to be the expected value of g over that level set is "the best" we can do.

Since the constant function $\mathbf{1}$ is in $\mathcal{F}$ (by Remark 2.2) and the MC condition holds for all $f \in \mathcal{F}$, it follows that

$$\left| v_p - v \right| = \left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_v} [g(x)] - v \right| = \left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_v} [g(x) - v] \right| \leq \frac{\epsilon}{2}$$

for all X_v such that $\Pr_{x\sim\mathcal{D}}[x \in X_v] \geq \gamma$. Therefore, v can be at most $\epsilon/2$ away from v_p. Hence, when we interchange $h(x) = v$ for $h(x) = v_p$, we obtain that

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_v} [f(x) \cdot (g(x) - v_p)] \right| \leq \left| \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_v} [f(x) \cdot (g(x) - v)] \right| + \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_v} \left| f(x) \cdot (v - v_p) \right| \leq \epsilon$$

for all X_v such that $\Pr_{x\sim\mathcal{D}}[x \in X_v] \geq \gamma$. Therefore, we have concluded that

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

for all X_v such that $\Pr_{x\sim\mathcal{D}}[x \in X_v] \geq \gamma$. By Definition 2.25 this means that $\mathcal{P} := \{X_v\}$ is an $(\mathcal{F}, \epsilon, \gamma)$-approximately multicalibrated partition of $\mathcal{X}$ for g on $\mathcal{D}$, as we wanted to show. □

Having introduced the notion of a multicalibrated partition, we know study the notion of the complexity of a partition, similar to how we considered the complexity of a multicalibrated predictor in Section 2.2.

Building on Definition 2.14, we introduce the class $\mathcal{F}_{t,q,k}$ of partitions:

Complexity of a partition. As we started to develop in Section 2.1, a key property of a multicalibrated partition is that it is a *low-complexity* partition of the domain $\mathcal{X}$. We now formalize this idea.

Definition 2.28. Given a set of functions $\mathcal{F} = \{f\}$ on a finite domain $\mathcal{X}$, $\mathcal{F}_{t,q,k}$ denotes the class of partitions $\mathcal{P}$ of $\mathcal{X}$ into k pieces $\mathcal{P} = (\{P_1, \ldots, P_k\})$ such that there exists $f_m \in \mathcal{F}_{t,q}$ satisfying $P_i = f_m^{-1}(i)$ for all $i \in [k]$. (Hence, $f_m \colon \mathcal{X} \to [k]$.)

We note that by writing $(\{P_1, \ldots, P_k\})$ we indicate that this is an ordered partition.

The condition $P_i = f^{-1}(i)$ stated in Definition 2.28 ensures that we are always able to know to which level set each $x \in \mathcal{X}$ belongs to by performing an oracle call to a function in $\mathcal{F}_{t,q}$. Intuitively, we are enumerating each $P \in \mathcal{P}$ with an integer in $[k]$, and then Definition 2.28 requires the existence of a function in $F_{t,q}$ that we use to query in which P each $x \in \mathcal{X}$ belongs to. We call this f the *partition membership function*. Moreover, this f is constant on each $P \in \mathcal{P}$.

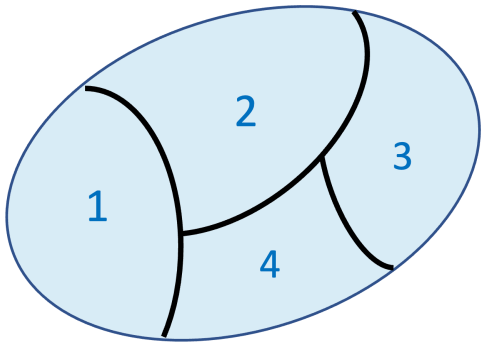

Figure 2.5: Illustration of the partition membership function. For example, if $x \in P_3$, then $f_m(x) = 3$. Thus, $P_3 = f_m^{-1}(3)$.

Having formalized the complexity class $\mathcal{F}_{t,q,k}$ of partitions, we can finally state the theorem that will be the backbone of all our results in Part II:

Theorem 2.29. *Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a class of functions $f \colon \mathcal{X} \to [0,1]$, $g \colon \mathcal{X} \to [0,1]$ an arbitrary function, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, and $\epsilon, \gamma > 0$. Then there exists an $(\mathcal{F}, \epsilon, \gamma)$-approximately multicalibrated partition $\mathcal{P}$ of $\mathcal{X}$ for g on $\mathcal{D}$ such that $\mathcal{P} \in \mathcal{F}_{t,q,k}$, where*

1. $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$,
2. $q = O(1/\epsilon)$,

3. $k = O(1/\epsilon)$.

Proof. We begin by invoking Theorem 2.19 with $\epsilon/2$ to obtain an $(\mathcal{F}, \epsilon/2, \gamma)$-approximately MC predictor $h\colon \mathcal{X} \to [0,1]$ that satisfies $h \in \mathcal{F}_{t,q}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$ and $q = O(1/\epsilon)$. By Claim 2.27, this yields an $(\mathcal{F}, \epsilon, \gamma)$-approximately MC partition $\mathcal{P}$ of $\mathcal{X}$. By Claim 2.11 and Remark 2.12, $|\mathcal{P}| = O(1/\epsilon)$ given that the predictor h has $O(1/\epsilon)$ level sets, and by the construction in the proof of Claim 2.27, the level sets of h and the $P \in \mathcal{P}$ are in 1-to-1 correspondence. Hence, $k = O(1/\epsilon)$. Lastly, given the 1-to-1 correspondence between the level sets of h and the sets $P \in \mathcal{P}$, it is clear that there exists a function $f_m \in \mathcal{F}_{t,q}$ satisfying $P_i = f_m^{-1}(i)$ for all $i \in [k]$. We construct this f_m by using the predictor h, which we know belongs to the complexity class $\mathcal{F}_{t,q}$. Namely, we define f_m as follows: let $v_1, \ldots, v_k$ be an arbitrary ordering of the k output values of h. Then, f_m maps each v_i to its index i in this ordering. We claim that $f_m \in \mathcal{F}_{t,q}$. Indeed, let C_h be the oracle-aided circuit that computes h. It is enough that we hard-wire the values $i \in [k]$ as determined by the mapping that we just defined, which we describe using a look-up table. Hence, the circuit that computes f_m only has an additional k number of gates [Bar22, §9.1.1.], but since $k = O(1/\epsilon)$ it follows that $f_m \in \mathcal{F}_{t,q}$, since the term $O(1/\epsilon)$ is absorbed by the parameter t. □

2.4 Bounding the computational complexity

While for the purposes of this thesis we are only concerned with the complexity of h relative to $\mathcal{F}$ (i.e., as defined in Definition 2.13), in this section we clarify how this relates to other notions of complexity about h present in the literature. In other words, in Algorithm 1 we are assuming that Step 3 in the pseudocode (namely, searching over the space $\mathcal{F}$ of distinguishers to find some f for which the MA/MC condition is violated) takes constant-time. While this is the right approach when considering the complexity of h with respect to $\mathcal{F}$, we cannot ignore this runtime when learning such a predictor h in practice.

The original paper of [HKRR18], besides considering the complexity of h relative to $\mathcal{F}$, also considers two other notions of complexity: one is the *complexity of learning*, and the other is the *sample complexity*. The notion of relative complexity considered in Theorem 2.19 only considers that size of the circuit that implements the simulator h. But from the perspective of learning theory, one should also consider how such a circuit can be *learnt*. In that case, as one might expect, [HKRR18, Section 3] show that the running time scales linearly with $|\mathcal{F}|$ (i.e., the time to iterate through each $f \in \mathcal{F}$) and with t, where t is an upper bound on the running time required to evaluate any $f \in \mathcal{F}$.

Hébert-Johnson et al. then study whether we can improve the linear dependence on $|\mathcal{F}|$ by exploiting structure within the collection of distinguishers $\mathcal{F}$. However, [HKRR18] show that weak agnostic learning of a class $\mathcal{F}$ is equivalent to learning an $(\epsilon, \mathcal{F})$-MC predictor up to polynomial factors; i.e., they show that the two problems reduce to each other. Therefore, in the positive direction, if there is an efficient (weak) agnostic learner for a class $\mathcal{F}$, then we can achieve efficient multicalibration with respect to $\mathcal{F}$. In the negative direction, this also means that learning an MC predictor for $\mathcal{F}$ is as hard as weak agnostic learning on $\mathcal{F}$. In particular, agnostic learning is known to be a notoriously hard problem in the learning theory literature (see, e.g., [KMV08; Fel09]).

In the context of learning theory, the predictor h should be learnt from labeled samples. In other words, it is not feasible in this context to assume full access to the domain $\mathcal{X}$, and instead the

learning algorithm only receives samples $(x, g(x))$. This makes sense, for example, in the context of machine learning, where a predictor receives a small set of labeled samples. The question then becomes: what is the least number of samples that are required to learn a multicalibrated predictor? This question is investigated thoroughly in [HKRR18].

However, for our purposes in this thesis, we are only concerned with the relative complexity of h with respect to $\mathcal{F}$, and we do not further discuss the computational complexity of learning.

2.5 Outcome Indistinguishability

To finish this chapter, we briefly summarize another important perspective on the notions of multiaccuracy and multicalibration. Recently, Dwork et al. established the framework of *outcome indistinguishability*, which studies questions in algorithmic fairness drawing on an understanding of computational indistinguishability developed in complexity theory and cryptography [DKR+21]. Dwork et al. inquire the meaning behind the concept of individual probabilities in algorithmic fairness. Namely, what does it mean that someone has a probability of a 5-year survival after a certain diagnosis, given that this is a non-repeatable event? Understanding this notion seems imperative in the case of algorithmic fairness: in this setting, the simulator h is mapping individuals in $\mathcal{X}$ to a score in $[0, 1]$, which is what is understood as an individual probability for some event.

Inspired by the literature on statistical forecasting [Daw85], Dwork et al. recently proposed the notion of *outcome indistinguishability* (OI): predictors that are OI yield a generative model for outcomes that cannot be efficiently refuted on the basis of the real-life observations produced by Nature. That is, the goal in the OI framework is to produce predictions that are indistinguishable from the ground truth. As it is the case in multiaccuracy and multicalibration, this notion captured through a class of distinguishers $\mathcal{F}$ and requiring indistinguishability between two predictors with respect to $\mathcal{F}$. They provide a hierarchy of four definitions for a predictor, based on how much access the distinguishers have to the predictor (in this setting, we think of the distinguishers as a board that is examining how good a predictor h is): no-access OI, sample-access OI, oracle-access OI, and code-access OI. They then show that their notion of no-access OI essentially corresponds to the notion of multiaccuracy, while their notion of sample-access OI essentially corresponds to the notion of multicalibration.

The OI framework and its connections to multiaccuracy and multicalibration has proven to be very useful in recent work, especially in the setting of learning theory and loss minimization [GHK+22; GHK+23; GKR23]. In particular, these works try to reconcile three different perspectives on the goal of learning: loss minimization, fairness (using MA and MC), and indistinguishability (using the OI framework).

3
Regularity Lemmas

Our result [...] appears to be the general result underlying the known connections between "regularity" results in graph theory, "decomposition" results in additive combinatorics, and the Hardcore Lemma in complexity theory.

Trevisan et al. [TTV09]

While the notion of multiaccuracy, and the proof that we can efficiently build a multiaccurate predictor using a boosting-type algorithm, recently emerged from the new field of algorithmic fairness, this notion turns out to be intimately related —in fact, equivalent— to the older *regularity lemma* of Trevisan, Tulsiani, and Vadhan [TTV09]. The regularity lemma of Trevisan et al. was published about 10 years before Hébert-Johnson et al. proposed the notion of multiaccuracy [HKRR18], and, crucially, has several deep connections to fundamental theorems in theoretical computer science and mathematics, as we later develop in this chapter. These connections are what establish the research question that we investigate in this thesis: Given that multicalibration is a stronger notion than multicaccuracy, how do these fundamental theorems change when we consider them through multicalibration rather than through multiaccuracy?

Formally, Trevisan et al. proved the following theorem:

Theorem 3.1 ([TTV09])**.** *Let $\mathcal{X}$ be a finite set, $\mathcal{D}$ a probability distribution over $\mathcal{X}$, $\mathcal{F}$ be a collection of functions $f\colon \mathcal{X} \to [0,1]$, $\epsilon > 0$ an approximation parameter, and $g\colon \mathcal{X} \to [0,1]$ an arbitrary bounded function.*

Then there is a function $h\colon \mathcal{X} \to [0,1]$ satisfying $\mathbb{E}_\mu[h] = \mathbb{E}_\mu[g]$ that is

1. *efficient relative to $\mathcal{F}$: h has complexity $\epsilon^{-O(1)}$ relative to $\mathcal{F}$, and*
2. *indistinguishable from g: for every $f \in \mathcal{F}$, we have*

$$\left|\mathbb{E}_{x\sim\mathcal{D}}[f(x)\cdot(g(x)-h(x))]\right| \leq \epsilon.$$

The notion of indistinguishability from g corresponds exactly to h being $(\mathcal{F},\epsilon)$-multiaccurate, as we noted in Chapter 2.

3.1 Complexity of the simulator

After the work of Trevisan et al. [TTV09], subsequent papers proved variants of the regularity lemma, mostly concerned with improving the efficiency of the simulator h with respect to the distinguishers f [JP14; CCL18; Skó15; Sko16; Skó16; VZ13]. The formulation used by Jetchev and Pietrzak [JP14] is suitable for cryptographic applications: indeed, their motivation for considering the regularity lemma is due to its applications to leakage-resilient cryptosystems. For this reason, [JP14] and [CCL18] refer to the regularity lemma as the *leakage simulation lemma*.

All of the proofs for the regularity lemma and its variations fall into two main proof techniques: a boosting proof via an energy-decrease argument (as discussed in Section 2.2, or a proof based on the min-max theorem, which we summarize in the next section. The min-max-type proofs also include several variations, such as the multiplicative weight update (MWU) method incorporating with KL-projections [VZ13]. The different proofs and the differences between parameters is discussed in the work of Chen et al. [CCL18].

Lower bounds. In [CCL18], they prove that the simulator *must* have a relative complexity of $q = \Omega(1/\epsilon^{-2})$ to the distinguisher family by establishing a black-box lower bound, where a simulator can only use the distinguishers in a black-box way. Given that the notion of multiaccuracy in algorithmic fairness corresponds to the regularity lemma, as we have just established, this lower bound shown in [CCL18] is also applicable to the construction of multiaccurate predictors, and it is thus also applicable to the construction of multicalibrated predictors, given that MA is a strictly weaker notion than MC.

3.2 Min-max proof

Another way to prove Theorem 3.1, as formalized in [TTV09], is by using Von Neumann's min-max theorem for two-player zero-sum games. The min-max theorem is also known as the *linear programming duality* or the *finite-dimensional Hahn-Banach Theorem*, and it has become a widely-used tool in theoretical computer science, and particularly in game theory [FS99]. We give a brief overview of how this argument works, without going into the technical details.

The min-max theorem applies to the setting of zero-sum games between two players. For every mixed strategy V (as a distribution over their strategy space $\mathcal{V}$) for Player 1, Player 2 has a response $W \in \mathcal{W}$ that guarantees $\mathbb{E}[F(V,W)] \geq 0$, where F can be an arbitrary function and is called the *payoff*. The min-max theorem states that there must exist a Player 2's mixed strategy W^* that guarantees $\mathbb{E}[F(V,W^*)] \geq 0$ for *all* strategies $V \in \mathcal{V}$ of Player 1. In other words, in any finite two-player zero-sum game, if for every distribution over the actions of Player 1 there exists some action for Player 2 that guarantees him an expected utility of v, then there exists some (universal) distribution of actions for Player 2 such that no matter what action Player 1 picks, Player 2 is still guaranteed an expected utility of v [VZ13]. The min-max theorem can be used to prove Impagliazzo's Hardcore Lemma [Imp95] and the Dense Model Theorem [RTTV08], among many other results.

In the setting of regularity/multiaccuracy, the two-player zero-sum game is defined as follows. Let $\mathcal{H}$ be the set of all bounded functions $h\colon \mathcal{X} \to [-1,1]$ that have complexity at most t with respect to $\mathcal{F}_t$, where $\mathcal{F}_t$ denotes the class of functions that have complexity at most t with respect

to $\mathcal{F}'$, which denotes the closure of $\mathcal{F}$ under negation; i.e., $\mathcal{F}' := \{f, -f \mid f \in \mathcal{F}\}$. The parameter t is set accordingly in the formal proof in [TTV09]. Then, the key idea is to set up the following game:

1. Player 1 picks a simulator h from $\mathcal{H}$.
2. Player 2 picks a distinguisher f from $\mathcal{F}'$.

The payoff function is defined as

$$\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[f(x)g(x) - f(x)h(x)].$$

Then, by applying the min-max theorem, [TTV09] show that there exists some $\overline{h} \in CH(\mathcal{H})$, where $CH(\mathcal{H})$ denotes the set of convex combinations of functions in $\mathcal{H}$, such that for all $f \in \mathcal{F}'$,

$$\mathop{\mathbb{E}}_{x\sim\mu}[f(x) \cdot (g(x) - \overline{h}(x))] \leq \epsilon/2.$$

Trevisan et al. then use the function $\overline{h}$ to build the simulator h that proves the regularity lemma (Theorem 3.1). The function $\overline{h}$ alone is not enough to prove it because $\overline{h} \in CH(\mathcal{H})$, and the convex combination may not have low-complexity.

Chen et al. [CCL18] and Vadhan and Zheng [VZ13] also provide a proof of the regularity lemma using the min-max theorem.

3.3 Structure and randomness in combinatorics

The regularity lemma of Trevisan et al. was being independently used in the field of combinatorics, where results of this sort are referred to as *decomposition* theorems, a term coined by Timothy Gowers [Gow10]. We believe that their casting of the regularity lemma provides a useful perspective for understanding the underlying principle behind the regularity lemma and its variants. The key idea behind these type of statements, as described by Terence Tao, is the following: In order to deal with a large object of unspecficied or unusable structure, we decompose it into more usable components [Tao07]. Usually, we decompose the object into a structured component, a pseudorandom component, and possibly an error term:

Object = Structured component + Pseudorandom component (+ error)

Tao calls this fundamental phenomenon a *dichotomy* between structure and pseudorandomness. Some examples of structured objects include complete bipartite graphs, functions with some periodicity, or objects with some algebraic structure. On the other hand, pseudorandom objects mimic the behaviour of random objects in some sense. The contribution of the pseudorandom and error components is shown to be negligible [Tao07].

In this light, the regularity lemma (Theorem 3.1) can be re-stated as follows [TTV09, Remark 1.4]: Given an arbitrary function g, we can find two functions $h_1 \colon \mathcal{X} \to [0, 1]$ and $h_2 \colon \mathcal{X} \to [-1, 1]$ such that $g = h_1 + h_2$, where h_1, h_2 satisfy that

1. h_1 has low complexity, and

2. h_2 is nearly orthogonal to all $f \in \mathcal{F}$; i.e., $|\langle h_2, f\rangle| \leq \epsilon$, where the inner product is defined as

$$\langle f, g\rangle := \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[f(x)g(x)].$$

As noted by [TTV09], this condition can be made nicer by introducing the norm

$$||g||_{\mathcal{F}} = \min_{f\in\mathcal{F}} \Big| \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[f(x)g(x)]\Big|,$$

in which case this Condition 2 becomes $||h_2||_{\mathcal{F}} \leq \epsilon$.

Then, we see how the Trevisan et al. regularity lemma fits exactly into Tao's framework: namely, the "object" corresponds to g (which can be arbitrarily complex), the "structured component" corresponds to $h_1 := h$ (since h has low complexity), and the "pseudorandom component" corresponds to $h_2 := g - h$.

Hence we can understand the regularity lemma as as part of this broader framework, where object is a superposition of a structured object and a pseudorandom error. For these type of structure theorems, the core idea behind their proofs is to use an iterative procedure which is based on this dichotomy. This iterative procedure is then shown to terminate using a potential energy argument. Indeed, this is exactly how we showed the multiaccuracy theorem in Chapter 2 (which corresponds to the regularity theorem). We can understand this iterative procedure as follows: At each iteration, if an object does not behave pseudorandomly, then it correlates with a nontrivial structured object, and we use this correlation to make the update. We continue until the initial object behaves pseudorandomly.

Other examples of this dichotomy include [Tao05; Tao07]: The spectral decomposition of a self-adjoint operator, Szemerédi's graph regularity lemma, orthogonal decomposition in Hilbert spaces, structure in Reed-Muller codes, among others. This dichotomy is also the key behind some fundamental results in additive combinatorics, including Szemerédi's theorem on arithmetic progressions and the Dense Model Theorem. The Dense Model Theorem in turn helped establish the famous Green-Tao theorem, which states that the primes contain arbitrarily long arithmetic progressions [GT08].

We briefly describe one of these examples; namely, orthogonal decomposition in Hilbert spaces. In the next section, we turn to Szemerédi's graph regularity lemma, which is one of the fundamental implications of the regularity lemma (Theorem 3.1).

3.3.1 An example: Orthogonal decomposition in Hilbert spaces

An example of the structure-pseudoranomdness dichotomy is as follows [Tao07]: we can show that we can decompose a vector f in a Hilbert space into its orthogonal projection f_{str} plus its orthogonal projection f_{psd} onto the orthogonal complement $V^{\perp}$ of V:

$$f = f_{str} + f_{psd}.$$

We now formalize this idea. Let H denote a real finite-dimensional Hilbert space, and let $S \subseteq H$ denote a known set S of "basic structured objects". We think of the dimension of H as being very large. We assume that $||v||_H \leq 1$ for all $v \in S$. The idea is that we want to efficiently represent

elements of H using S. Namely, vectors in S as seen as *structured*, whereas vectors that have small inner product to all vectors in S are seen as *pseudorandom*.

Definition 3.2. Given $f \in H$, we say that a vector f is (M, K)-*structured* for some $M, K > 0$ if there exists a decomposition

$$f = \sum_{1 \leq i \leq M} c_i v_i$$

with $v_i \in S$ and $c_i \in [-K, K]$ for all $1 \leq i \leq M$.

Definition 3.3. We say that f is ϵ-*pseudorandom* for some $\epsilon > 0$ if $|\langle f, v \rangle_H| \leq \epsilon$ for all $v \in S$.

We can now state the following structure theorem:

Theorem 3.4 (Non-orthogonal weak structure theorem [Tao07, Corollary 2.5])**.** *Let H, S be as above. Let $f \in H$ be such that $||f||_H \leq 1$, and let $0 < \epsilon \leq 1$. Then there exists a decomposition*

$$f = f_{str} + f_{psd},$$

where f_{str} is $(1/\epsilon^2, 1/\epsilon)$-structured and f_{psd} is ϵ-pseudorandom.

The proof follows an energy decrease argument using Cauchy-Schwarz and Pythagoras' theorem. At each step, $||f_{psd}||_H^2$ decreases by at least ϵ^2, and thus the algorithm terminates after at most $1/\epsilon^2$ such iterations.

3.4 Szemerédi regularity lemma

After having described the general framework surrounding Theorem 3.1, we now turn to Szemerédi's regularity lemma in graph theory, which is another canonical example of the dichotomy between structure and randomness. On a high-level, Szemerédi's regularity lemma states that large dense graphs can be decomposed into low-complexity partitions and regular graphs between partition classes. In Tao's framework, the "object" here corresponds to a graph which can be arbitrarily large and dense (hence complex), the "structured component" corresponds to the low-complexity partition classes, and the "pseudorandom component" corresponds to the regular graphs between partition classes.

There are two versions of this result: the original and stronger result, which is called *Szemerédi regularity lemma*, and the weaker version shown by Frieze and Kannan, which obtains better parameters for algorithmic applications [FK99]. In this section, we describe this fundamental regularity result in graph theory for two reasons: first, it provides another illustration of the type of decomposition theorems that we are describing in this chapter. Second, and more importantly, Trevisan et al. show that their regularity lemma (Theorem 3.1) yields the weak Szemerédi regularity lemma of Frieze and Kannan as a corollary. In this section, we will unpack this implication, and show how we can use a multiaccurate predictor to prove the Frieze-Kannan regularity lemma. This connection was recently explored in Dwork et al. [DLLT23] in the context of algorithmic fairness and in Skórski [Skó17] in the context of low-complexity approximations in cryptography. They both show, in different ways and using different terminology, that a multicalibrated predictor can be used to prove a stronger variant of the regularity of Frieze and Kannan. Dwork et al. call this notion *intermediate regularity* [DLLT23].

Describing these implications will help set-up the type of analysis that we will be repeatedly performing in Part II of this thesis in order to obtain our results. Namely, we begin by taking the regularity lemma of Trevisan et al. (Theorem 3.1) and we "translate" it to some precise context; in this case, in the area of graph theory. This entails instantiating the domain $\mathcal{X}$, the class of distinguishers $\mathcal{F}$, and function g accordingly. Then, by applying Theorem 3.1 (i.e., the multiaccuracy theorem/regularity lemma), we obtain a multiaccurate predictor h for g, which allows us to prove some other fundamental theorem. Having understood this proof, the next step is then to reproduce this same procedure but using the *multicalibrated* theorem instead. Then, by reproducing the first proof obtained with a multiaccurate predictor but using a multicalibrated predictor instead, we are able to obtain a stronger and more general version of the original fundamental theorem. In the case of graph theory, multicalibration allows us to obtain a stronger version of the Frieze-Kannan regularity lemma.

3.4.1 Frieze-Kannan regularity

Definition 3.5 (Density)**.** Let $G = (V, E)$ a graph, where V denotes the vertex set and $E \subseteq V \times V$ denotes the edge set. For disjoint sets $S, T \subseteq V$, let $e_G(S, T)$ denote the number of edges between S and T. The *density* $d_G(S, T)$ is defined as

$$d_G(S, T) = \frac{e_G(S, T)}{|S||T|}.$$

We will drop the subscripts of e_G and d_D if it is clear to which graph G we are referring to.

The following definition requires a global regularity guarantee on the partition of vertices of a graph:

Definition 3.6 (Frieze-Kannan ϵ-regularity)**.** A partition $\mathcal{P} = \{V_1, \ldots, V_m\}$ of the vertices V of a graph G satisfies *Frieze-Kannan ϵ-regularity* if

$$|e_G(S, T) - \sum_{j,k \in [m]} d_G(V_j, V_k)|S \cap V_j||T \cap V_k|| \leq \epsilon |V|^2$$

for all disjoint $S, T \subseteq V$.

Frieze and Kannan showed that the above definition is indeed achievable for any graph G [FK99]:

Theorem 3.7 (Frieze-Kannan Regularity Lemma [Skó17, Thm. 4])**.** *For every graph G there exists a partition $V_1, \ldots, V_k$ of the vertices V and real numbers $d_{i,j}$ such that*

$$\Big| \sum_{i,j} e_G(S \cap V_i, T \cap V_j) - \sum_{i,j} d_{i,j}|S \cap V_i||T \cap V_j| \Big| \leq \epsilon |V|^2$$

for all $S, T \subseteq V$. Moreover, the partition is generated by $O(\epsilon^{-2})$ subsets of V. In particular, k is at most $2^{O(\epsilon^{-2})}$.

3.4.2 Szemerédi regularity

A stronger regularity condition on a partition of the vertices is the following one. First, we say that a pair (X, Y) is regular if the density is approximately preserved:

Definition 3.8 (ϵ-regular)**.** We say that a disjoint pair $X, Y \subseteq V$ in a graph G is *ϵ-regular* if for every $S \subseteq X$ such that $|S| \geq \epsilon|X|$ and $T \subseteq Y$ such that $|T| \geq \epsilon|Y|$ we have

$$|d_G(S,T) - d_G(X,Y)| \leq \epsilon.$$

That is, a disjoint pair (X, Y) is ϵ-regular if it is distributed pseudorandomly. The natural question is then: Can we find a partition of the vertices of the graph such that most parts are ϵ-regular? Formally:

Definition 3.9 (Szemerédi ϵ-regularity)**.** A partition $\mathcal{P} = \{V_1, \ldots, V_m\}$ of the set of vertices V of a graph G satisfies *Szemerédi ϵ-regularity* if

$$\sum_{\substack{j,k \in [m] \\ (V_j, V_k) \text{ not } \epsilon\text{-regular}}} |V_j||V_k| \leq \epsilon |V|^2.$$

Even for this stronger notion, it is also possible to achieve it (although with worse parameters than in the case of Frieze-Kannan, as one would expect):

Theorem 3.10 (Szemerédi Regularity Lemma, variant 1 [Skó17, Thm. 1])**.** *For every graph G, there exists a partition $V_1, \ldots, V_k$ of vertices such that for all up to ϵ-fraction of the pairs (i, j),*

$$\left| e({}_G S, T) - d_G(V_i, V_j)|S||T| \right| \leq \epsilon |V_i||V_j|$$

for any $S \subseteq V_i$, $T \subseteq V_j$ such that $|S| \geq \epsilon|V_i|$, $|T| \geq \epsilon|V_j|$. Moreover, the size of the partition is at most a power of twos of height $O(\epsilon^{-2})$.

Szemerédi's Regularity Lemma can be shown using the same type of iterative argument that we have described in Chapters 2 and 3: Namely, we begin with an arbitrary partition of the graph. While the partition is not ϵ-regular, we find the subsets S and T which "witness" this irregularity, and we refine the partition using these subsets. Then, we use an potential energy decrease argument to argue that this procedure terminates after some bounded number of steps. We see that this corresponds exactly to Algorithm 1 in Chapter 2 and the subsequent analysis showing how to build a multiaccurate predictor. Namely, we begin with a trivial predictor h. While the graph is not multiaccurate, we find some distinguisher who "distinguishes", and we update h using this distinguisher. Then, we also use a potential energy decrease argument (see proof of Theorem 2.17).

For many years, it was not known whether the tower-type bound stated in Theorem 3.10 (namely, a power of twos of height poly$(1/\epsilon)$) was unavoidable. In 1997, Timothy Gowers showed that this is indeed the case [Gow97].

3.5 Multiaccuracy corresponds to Frieze-Kannan weak regularity

The fact that multiaccuracy (equivalently, the regularity lemma of [TTV09]) corresponds to Frieze-Kannan regularity was originally observed in [TTV09], and further explored in [Skó17; DLLT23]. To show this correspondence, given a graph $G = (V, E)$ we instantiate the regularity lemma of Trevisan et al. (Theorem 3.1) with the appropriate domain $\mathcal{X}$, class of distinguishers $\mathcal{F}$, and function g:

- We define the domain $\mathcal{X}$ as the set of edges in a complete graph of V. That is,

$$\mathcal{X} = \{(a, b) \mid a, b \in V\}.$$

- In this way, we can see the graph G as defining a boolean function $g\colon \mathcal{X} \to \{0, 1\}$.
- Lastly, the set of distinguishers $\mathcal{F}$ is instantiated as follows. For every two disjoint set of vertices S, T, we let the function $f_{S,T}\colon \mathcal{X} \to \{0, 1\}$ be defined as the characteristic function of the set of edges having one endpoint in S and one in T. Then $\mathcal{F}$ contains all $f_{S,T}$ for every $S, T \subseteq V$.

In order to understand this instantiation of $\mathcal{X}, g$, and $\mathcal{F}$, we provide the following example. Consider the following graph G:

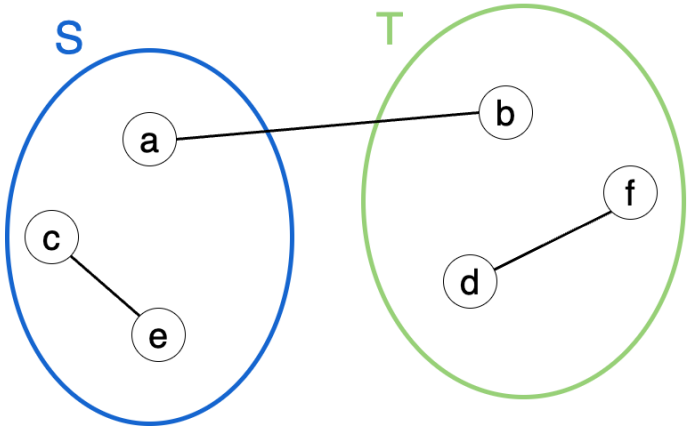

Figure 3.1: Example graph to illustrate the correspondence between Frieze-Kannan regularity and multiaccuracy.

In this example, we see that $\mathcal{X}$ corresponds to

$$\mathcal{X} = \{(a,b),(a,c),(a,d),(a,e),(a,f),(b,c),(b,d),(b,e),(b,f),(c,d),(c,e),(c,f),(d,e),(d,f),(e,f)\}.$$

The function g returns 1 in the following cases:

$$g((a,b)) = 1, \qquad g((c,e)) = 1, \qquad g((d,f)) = 1,$$

and returns 0 for the rest of values in $\mathcal{X}$. Essentially, if we consider the adjacency matrix of G, the function g returns 1 on a pair of vertices if and only the corresponding value of that pair in the adjacency matrix of G corresponds to 1.

Lastly, given the sets S and T of vertices illustrated in Figure 3.1, it follows that f returns 1 in the following cases:

$$\begin{aligned}
&f_{S,T}((a,b)) = 1, \qquad f_{S,T}((a,d)) = 1, \qquad f_{S,T}((a,f)) = 1,\\
&f_{S,T}((c,b)) = 1, \qquad f_{S,T}((c,d)) = 1, \qquad f_{S,T}((c,f)) = 1,\\
&f_{S,T}((e,b)) = 1, \qquad f_{S,T}((e,d)) = 1, \qquad f_{S,T}((e,f)) = 1,
\end{aligned}$$

and returns 0 for the other values in $\mathcal{X}$. That is, $f_{S,T}$ only returns 1 on an edge that crosses between S and T. (This edge might not actually appear in the graph G, since $f_{S,T}$ does not depend on g.)

By applying the multiaccuracy theorem to $\mathcal{X}, g$, and $\mathcal{F}$ (Theorem 2.17), we obtain a simulator $h\colon \mathcal{X} \to [0,1]$ that satisfies the MA guarantee. We now justify that we can see h as a weighted graph H that approximates G as required by the Frieze-Kannan Weak Regularity Lemma [TTV09]. That is, we can see h as defining a sort of adjacency matrix for the graph G, except the matrix values are allowed to be real-valued, unlike in the case of proper adjacency matrices, which only take boolean values.

By the MA condition, we know that for all $f_{S,T} \in \mathcal{F}$,

$$|\,\mathbb{E}[f_{S,T}(x)g(x)] - \mathbb{E}[f_{S,T}(x)h(x)]| \leq \epsilon.$$

In order to ease notation, we state the condition for $x \in E$, unlike for $(a,b) \in E$ where $a, b \in V$, as we did it in the example above. As we just illustrated, $f_{S,T}$ is a "cut function": it only return 1 when evaluated on an edge that crosses the cut (with respect to the vertex sets S and T). By the definition of g, we also know that $g(x) = 1$ if and only if $x \in E$; that is, if x is an edge in G.

Let $\mathcal{P} = \{V_1, \ldots, V_m\}$ be the partition induced by the multiaccurate predictor h. Namely, we construct this partition from the function h by taking all possible intersections of the sets S_i, T_i and their complements, such that h is constant on the edges between each pair of parts. This explains why the number of pieces in the partition stated in the Frieze-Kannan regularity lemma (Theorem 3.7) is $2^{O(\epsilon^{-2})}$. Namely, by the Trevisan et al. regularity lemma (Theorem 3.1), we know that h can be described as a function of at most $k = \text{poly}(1/\epsilon)$ functions f_{S_i,T_i}. Hence, by the construction of $\mathcal{P}$ from h that we just described, it follows that $|\mathcal{P}| \leq 2^{2k}$.

Then, by the definitions of g, $f_{S,T}$, and h, it follows that

$$\mathbb{E}[f_{S,T}(x)g(x)] = \frac{\sum_{j,k\in[m]} e(S \cap V_j, T \cap V_k)}{|V|^2},$$

$$\mathbb{E}[f_{S,T}(x)h(x)] = \frac{\sum_{j,k\in[m]} d(V_j, V_k)|S \cap V_j||T \cap V_k|}{|V|^2}.$$

Then, by the MA guarantee, it follows that

$$\left| e(S,T) - \sum_{j,k\in[m]} d(V_j, V_j)|S \cap V_j||T \cap V_k| \right| \leq \epsilon |V|^2,$$

given that $\sum_{j,k\in[m]} e(S \cap V_j, T \cap V_k) = e(S,T)$. This corresponds exactly to the definition of Frieze-Kannan ϵ-regularity, and hence we have proved the Frieze-Kannan regularity lemma (Theorem 3.7) using a multiaccurate predictor h.

Using multicalibration instead of multiaccuracy. Having established the equivalence between multiaccuracy and Frieze-Kannan regularity, Dwork et al. study what type of graph regularity lemma we would obtain if we started with a multicalibrated predictor instead. Namely, they show that a variant of approximate multicalibration (which they define using the framework of outcome indistinguishability) gives rise to a graph partition that satisfies what they call *intermediate regularity*. Formally, this term is defined as follows:

Definition 3.11 ([DLLT23, Def. 6.6])**.** Let $X, Y, S, T \subseteq V$ in a graph G. We say that the pair (X, Y) is (S, T, ϵ)-*regular* if

$$|d_G(S \cap X, T \cap Y) - d_G(X, Y)| \leq \epsilon.$$

Definition 3.12 (Intermediate ϵ-regularity [DLLT23, Def. 6.7])**.** A partition $\mathcal{P} = \{V_1, \ldots, V_m\}$ of the set of vertices V of a graph G satisfies *intermediate ϵ-regularity* if

$$\sum_{\substack{j,k \in [m] \\ (V_j, V_k) \text{ not } (S,T,\epsilon)\text{-regular}}} |V_j||V_k| \leq \epsilon |V|^2.$$

The notion of intermediate regularity is strictly between the Frieze-Kannan regularity notion and Szemerédi's regularity notion. In [DLLT23], following a similar approach as [Skó17], they also show that we can prove the Szemerédi regularity lemma using a stronger multicalibration notion than approximate multicalibration plus some extra structural conditions on the set of vertices. In particular, this extra structural condition is stringent enough that this equivalence does not contradict Gower's result on the lower bound on the size of a partition that satisfies Szemerédi regularity (namely, a power of twos of height $O(\epsilon^{-2})$).

3.6 Implications of a multiaccurate predictor

In Section 3.4, we have seen how we can prove the Frieze-Kannan regularity lemma in graph theory using a multiaccurate predictor. But this is not the only fundamental theorem that can be derived from the regularity lemma of Trevisan et al. (Theorem 3.1): There are many other fundamental theorems in various areas of theoretical computer science that can be derived as corollaries of the regularity lemma. Some of these fundamental theorems include:

Impagliazzo's Hardcore Lemma (IHCL). Impagliazzo's Hardcore Lemma (IHCL) is a fundamental result in complexity theory that was first proved in 1995 [Imp95]. Informally, it states that if a function is somewhat hard to compute on average by a family of boolean functions (what we have been calling the *distinguishers*), then there is a large-enough subset of the inputs (called the "hardcore set") for which the function is very hard to compute, in the sense that g behaves like a random function in the eyes of the distinguishers.

Characterizations of pseudoentropy. Vadhan and Zheng showed that we can characterize pseudoentropy, which is a notion from information theory, in terms of hardness of sampling [VZ13]. This characterization yields a simpler construction of pseudorandom generators from one-way functions, among other applications.

The Dense Model Theorem (DMT). This is a result from additive combinatorics which states the following [RTTV08; GT08]. If we have a pseudorandom set R in a domain $\mathcal{X}$ (which can be very sparse) and a set D contained in R such that D occupies a large enough fraction of the space inside R, then there exists a model set M in $\mathcal{X}$ such that M occupies a large enough fraction of the space inside X and such that R and M are indistinguishable with respect to the class of distinguishers $\mathcal{F}$. The Dense Model Theorem is one of the crucial proof components used in Green and Tao's famous result that there exist arbitrarily long arithmetic progressions of primes [GT08].

Other implications of the regularity lemma of Trevisan et al. include applications in leakage resilient cryptopgrahy [JP14; CCL18], weak notions of zero-knowledge [CLP15], Yao's XOR theorem [GNW11], chain rules for computational entropy [GW11; JP14], and Chang's inequality in Fourier analysis of boolean functions [IMR14].

Our research question. Given the correspondence between building a multiaccurate predictor and the regularity simulation lemma (i.e., between Theorem 2.17 and Theorem 3.1), a natural question arises: How do all these implications generalize when we start from a multicalibrated predictor instead of from a multiaccurate predictor? This should yield more general and stronger theorems, given that multicalibration is a stronger notion that multiaccuracy. In Section 3.4, we saw the first instance of this approach being successful; namely, multicalibration yields a stronger notion of graph regularity than that of Frieze-Kannan. In this thesis, we focus on three of the implications the regularity lemma: Impagliazzo's Hardcore Lemma (Chapter 4, characterizations of pseudoentropy (Chapter 5), and the Dense Model Theorem (Chapter 6). We find a stronger, more general theorem in all three cases; moreover, the three of them present a parallel structure (which we summarize in Chapter 7). Therefore, by using the tools that have recently been developed in the field of algorithmic fairness (namely, the construction of multicalibration) and casting them back to the field of computational complexity, we are able to obtain stronger and more general versions of fundamental theorems that have been well-known for years.

Throughout the thesis, we will use ++ to denote the strengthened and more general versions of all theorems.

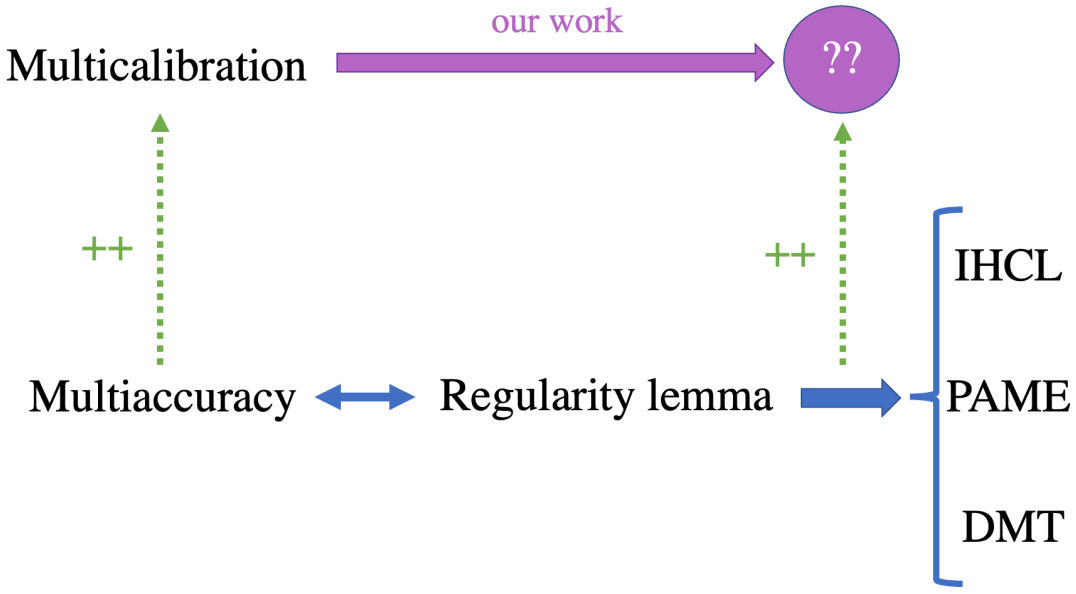

Figure 3.2: Diagram representing the research question underlying this thesis.

II

Main Theorems

4

Impagliazzo Hardcore Lemma

Consider a decision problem that cannot be $1-\delta$ approximated by circuits of a given size in the sense that any such circuit fails to give the correct answer on at least a δ fraction of instances. We show that for any such problem there is a specific "hard-core" set of inputs which is at least a δ fraction of all inputs and on which no circuit of a slightly smaller size can get even a small advantage over a random guess.

Russell Impagliazzo [Imp95]

Impagliazzo's Hardcore Lemma (IHCL) is a fundamental result in complexity theory that was first proved in 1995 [Imp95]. Informally, it states that if a function is somewhat hard to compute on average by a family of boolean functions (what we have been calling the *distinguishers*), then there is a large-enough subset of the inputs (called the "hardcore set") for which the function is very hard to compute, in the sense that g behaves like a random function in the eyes of the distinguishers. When we state the theorem formally, we will see that the "somewhat hard" assumption and the "very hard" conclusion are with respect to different families of distinguishers. In particular, the former family is an enlarged class of distinguishers than the latter family. In the case of IHCL, the distinguishers will correspond to circuits.

This is a type of hardness amplification from computational complexity [KS03]: IHCL is stating that from a boolean function g which is "mildly inapproximable" by circuits of some size, we can find a set of the inputs where g is "highly inapproximable" by circuits of slightly smaller size. We think of "mildly inapproximable" as saying that no circuit agrees with g on a fraction of inputs very close to 1, whereas we think of "highly inapproximable" as saying that no circuit can agree with g on a fraction of inputs larger than $1/2$. Moreover, we require the set of inputs in which g is highly inapproximable to have noticeable density.

In fact, Impagliazzo's Hardcore Lemma is deeply related to the notion of boosting in learning theory, as it was first shown in [KS03]. Namely, the boosting algorithm by Schapire and Freund [FS99] which converts weak learners into strong learns can be seen as the "opposite" of Impagliazzo's Hardcore Lemma, in the sense that boosting constructs a hypothesis which closely approximates a function whereas Impagliazzo's Hardcore Lemma proves that certain functions are hard to approxi-

mate [KS03]. Impagliazzo has also recently studied the connections between boosting the Hardcore Lemma [Lee17].

4.1 Definitions and the original IHCL statement

Before we state the theorem formally, we introduce the required definitions.

Definition 4.1 ($\mathcal{U}_\mathcal{X}$)**.** Given a domain $\mathcal{X}$, we use $\mathcal{U}_\mathcal{X}$ to denote the uniform distribution over $\mathcal{X}$. That is, every $x \in \mathcal{X}$ is assigned the same probability mass by $\mathcal{U}_\mathcal{X}$; namely, $1/|\mathcal{X}|$.

Definition 4.2 (δ-dense distribution)**.** A distribution A is *δ-dense* in a distribution B if for all $x \in \mathcal{X}$,

$$\delta \cdot \Pr[A = x] \leq \Pr[B = x].$$

If B is the uniform distribution on X, then this becomes

$$\delta \cdot \Pr[A = x] \leq \frac{1}{|\mathcal{X}|}.$$

In some settings, we will use *measures* instead of *distributions*:

Definition 4.3 (Measure)**.** A function $\mu \colon \mathcal{X} \to [0, 1]$ is a *measure* on $\mathcal{X}$.

The difference between a probability distribution and a measure is that a measure is not necessarily normalized. Still, a measure induces a probability distribution if we normalize it:

Definition 4.4 (From a measure to a distribution)**.** Given a measure $\mu \colon \mathcal{X} \to [0, 1]$, μ induces the probability distribution

$$\mathcal{D}_\mu(x) = \frac{\mu(x)}{\sum_{z \in \mathcal{X}} \mu(z)}.$$

Definition 4.5 (Density of a measure)**.** Given a measure $\mu \colon \mathcal{X} \to [0, 1]$, the *density* of μ, denoted $d(\mu)$, is defined as

$$d(\mu) = \mathop{\mathbb{E}}_{x \sim \mathcal{X}}[\mu(x)] = \frac{1}{|\mathcal{X}|} \cdot \sum_{x \in \mathcal{X}} \mu(x).$$

We say that μ is *δ-dense* if $d(\mu) \geq \delta$.

We remark that Definition 4.5 is exactly equivalent to Definition 4.2 if we scale μ so that its largest value is 1 and consider the distribution induced by this scaled μ.

When working with Impagliazzo's Hardcore Lemma, it is useful to translate between *measures* and *sets*. Historically, proofs involving the Hardcore Lemma always find a hardcore measure or distribution, but in applications it is often more intuitive to deal with sets instead of measures [Imp95; KS03; Hol05; TTV09]. This will also be true for us: we will prove our IHCL++ using distributions, but we will then state the corresponding version using sets, which will allow us to visualize the theorems pictorially.

We present the formal conversion between measures and sets in Section 4.3, but we state the definitions here because they provide some intuition on why we understand Definition 4.5 as a density measure.

The natural definition for the definition of a set is the following:

Definition 4.6 (δ-dense set)**.** Given a set $S \subseteq \mathcal{X}$, we say that S is *δ-dense* in $\mathcal{X}$ if $|S| \geq \delta \cdot |\mathcal{X}|$.

In other words, S is δ-dense in $\mathcal{X}$ if S occupies at least a fraction δ of the domain. Returning to our discussion on the size of the level sets in Chapter 2, recall that we called $\eta_p := |P|/|\mathcal{X}|$ the *density parameter* of the set P in $\mathcal{X}$. Hence, saying that $P \subseteq \mathcal{X}$ is δ-dense in $\mathcal{X}$ is equivalent to stating that $\eta_p \geq \delta$.

The idea behind how to build a set S from a measure μ so that the notion of density is preserved is to include an element $x \in \mathcal{X}$ into S with probability $\mu(x)$ (independently for each x). As we formalize in Section 4.3, this non-constructive construction of S ensures that if μ has noticeable density, then S has noticeable density as well with high probability.

Conversely, every $S \subseteq \mathcal{X}$ has a corresponding measure given by the associated characteristic function χ_S. That is, $\chi_S(x) = 1$ if $x \in S$, and $\chi_S(x) = 0$ if $x \notin S$. Then, we see that

$$|S|/|\mathcal{X}| = \Pr[x \in S] = \frac{1}{|\mathcal{X}|} \cdot \sum_{x \in \mathcal{X}} \chi_S(x) = \mathop{\mathbb{E}}_{x \sim \mathcal{X}}[\chi_S(x)],$$

which explains why the density of a measure is defined as in Definition 4.5.

Lastly, for the IHCL statement, we will need the following two definitions. In the case of IHCL, we will only work with boolean distinguishers; i.e., all f map from $\mathcal{X}$ to $\{0,1\}$. Similarly, in this chapter, the arbitrary function g will also always be boolean.

Definition 4.7 (δ-weakly hard)**.** Given a class $\mathcal{F}$ of functions $f\colon \mathcal{X} \to \{0,1\}$, a distribution $\mathcal{D}$ on $\mathcal{X}$, an arbitrary function $g\colon \mathcal{X} \to \{0,1\}$, and $\delta > 0$, we say that g is *δ-weakly hard* with respect to $\mathcal{F}$ on $\mathcal{D}$ if, for all $f \in \mathcal{F}$,

$$\Pr_{x \sim \mathcal{D}}[f(x) = g(x)] \leq 1 - \delta.$$

for all $f \in \mathcal{F}$. Alternatively, we say that g is *$(\mathcal{F}, \delta)$-weakly hard* on $\mathcal{D}$.

That is, all of the distinguishers $f \in \mathcal{F}$ fail to compute g on at least a δ fraction of the inputs $x \in \mathcal{X}$. As usual, if $\mathcal{D}$ is not specified, then we are implicitly working with the uniform distribution on the domain.

Definition 4.8 (Strongly hard & Hardcore distribution)**.** Given a class $\mathcal{F}$ of functions $f\colon \mathcal{X} \to \{0,1\}$, a distribution $\mathcal{D}$ on $\mathcal{X}$, an arbitrary function $g\colon \mathcal{X} \to \{0,1\}$, and $\epsilon > 0$, we say that g is *ϵ-strongly hard* with respect to $\mathcal{F}$ and a distribution $\mathcal{D}$ on $\mathcal{X}$ if, for all $f \in \mathcal{F}$,

$$\Pr_{x \sim \mathcal{D}}[f(x) = g(x)] \leq 1/2 + \epsilon.$$

In this case, we say that g is *$(\mathcal{F}, \epsilon)$-strongly hard* on $\mathcal{D}$, or that $\mathcal{D}$ is an *$(\mathcal{F}, \epsilon)$-hardcore* distribution for g.

That is, all of the distinguishers $f \in \mathcal{F}$ fail to compute g on about half of the inputs in the domain. We call this "strongly hard" because the fact that $\Pr_{x \sim X}[f(x) = g(x)] \leq 1/2 + \epsilon$ indicates that g is essentially behaving like a random function. That is, if $g \sim \text{Bern}(1/2)$, then on each $x \in \mathcal{X}$, $g(x)$ returns 0 with probability $1/2$, and 1 with probability $1/2$. Then, we expect that $f(x)$ will only match $g(x)$ on approximately half of the points in the domain, exactly as in Definition 4.8. We will explore this idea further in Section 4.4. We will be using the letter $\mathcal{H}$ to denote a hardcore distribution (and later the letter S to denote a hardcore set).

Clearly, Definition 4.7 (δ-weak hardness) is a weaker notion than Definition 4.8 when $\delta, \epsilon < 1/4$. What about the converse? Can we find a subset of inputs over which δ-weak hardness implies

strong hardness? Impagliazzo's Hardcore Lemma answers precisely this question:

Theorem 4.9 (IHCL, [Imp95; Hol05]). *Let $\mathcal{F}$ be a family of functions from a finite domain $\mathcal{X}$ to $\{0,1\}$ and $\epsilon, \delta > 0$. Then there exists an $s = \text{poly}(1/\epsilon, 1/\delta)$ such that if $g\colon \mathcal{X} \to \{0,1\}$ is a function, which for all functions $f_0\colon \mathcal{X} \to \{0,1\}$ having complexity at most s with respect to $\mathcal{F}$ satisfies*

$$\Pr_{x\sim X}[f_0(x) = g(x)] \leq 1 - \delta \qquad (g \text{ is } (\mathcal{F}_s, \delta)\text{-weakly hard}),$$

then there is a distribution $\mathcal{H}$ that is 2δ-dense in $\mathcal{U}_\mathcal{X}$ and for which

$$\forall f \in \mathcal{F}, \quad \Pr_{x\sim\mathcal{H}}[f(x) = g(x)] \leq 1/2 + \epsilon \qquad (g \text{ is } (\mathcal{F}, \epsilon)\text{-strongly hard on } \mathcal{H}).$$

Some considerations regarding the IHCL statement. We make some remarks about Theorem 4.9. First, the weakly-hardness assumption on g is with respect to the uniform distribution on $\mathcal{X}$, given that we write $\Pr_{x\sim\mathcal{X}}$. (We remark that this can be replaced for an arbitrary distribution, as we also discuss in Chapter 5. However, for the purposes of this chapter, we will only require weak hardness with respect to $\mathcal{U}_\mathcal{X}$.) For the strong hardness conclusion, we are sampling according to the distribution $\mathcal{H}$. There is another key difference between the assumption and the conclusion: the δ-weakly hard condition is with respect to a slightly class of distinguishers; namely, with respect to $\mathcal{F}_s$, which corresponds to the set of functions that have complexity at most s with respect to the functions in $\mathcal{F}$ (see Definition 2.14 and Remark 2.15).

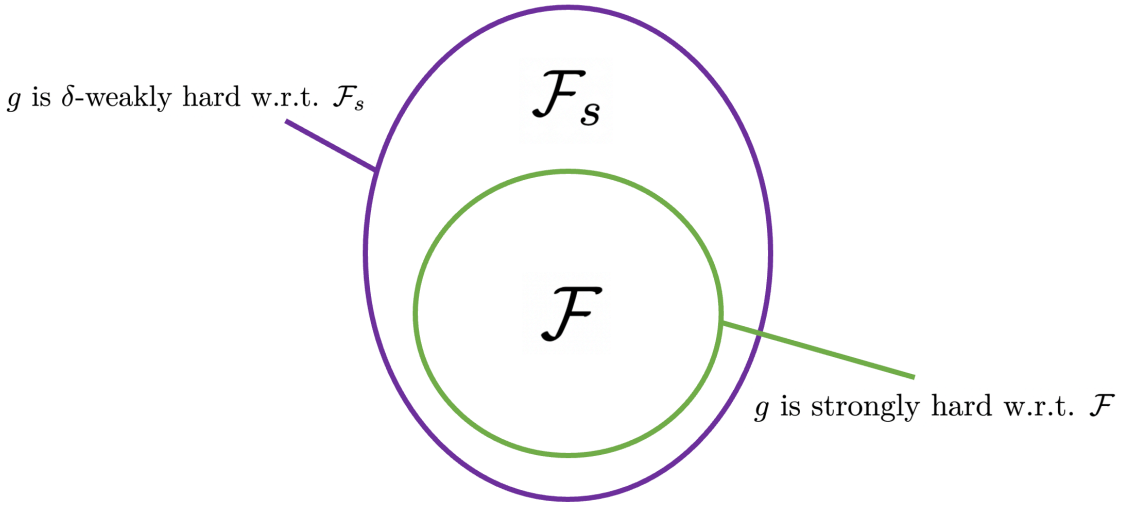

Figure 4.1: Illustration of the two classes of distinguishers considered in the IHCL statement (Theorem 4.9).

An important consideration about IHCL which will be key to our proposed IHCL++ is that the statement is proving two different things about the distribution $\mathcal{H}$:

- **Density.** The distribution $\mathcal{H}$ is 2δ-dense in $\mathcal{U}_\mathcal{X}$.
- **Indistinguishability.** When we sample according to $\mathcal{H}$, g is $(\mathcal{F}, \epsilon)$-strongly hard. We call

this the "indistinguishability" condition because, as we explained above, this corresponds to stating that g behaves like a random function with respect to the class of distinguishers $\mathcal{F}$.

Lastly, there are some important historical considerations regarding the density parameter of the distribution $\mathcal{H}$ in the IHCL statement. The original 1995 paper by Russell Impagliazzo proved the IHCL statement in two different ways. The first is a boosting proof via an energy decrease argument, similar to the boosting proof that we presented in Chapter 3. The second (due to Nisan) is a proof using von Neumann's min-max theorem, similar the the min-max proof of the regularity lemma that we summarized in Section 3.2.

However, the original theorem shown by Impagliazzo finds a hardcore measure of density δ, rather than 2δ. This difference is important because 2δ is the *optimal* density parameter for the hardcore measure. This is because if there exists a hardcore measure for g of density ρ, then g is $\big(\rho(1/2-\epsilon)\big)$-weakly hard on average on $\mathcal{U}_{\mathcal{X}}$ with respect to $\mathcal{F}$. It took 10 years for Holstein to prove that we can indeed achieve the 2δ density parameter [Hol05]. However, Trevisan et al.'s proof of IHCL using the regularity lemma (Theorem 3.1) is only able to recover the original δ-density parameter, but not Holstein's optimal 2δ-density parameter [TTV09]. This is a very important fact regarding the presentation of our results, because by adapting the proof of Trevisan et al. in order to use a multicalibrated predictor instead of a multiaccurate one, we are able to recover the original 2δ density parameter from IHCL. That is, a multiaccurate predictor does not seem to imply IHCL with optimal density parameters, but a multicalibrated predictor can. We also provide a second proof to IHCL++ different from [TTV09] which also obtains the optimal 2δ density parameter. Still, we believe that it is valuable to present the two proofs because they use different techniques and exploit different properties given by a multicalibrated predictor, and hence provide different insights into how the notion of multicalibration can provide stronger results related to the implications of the regularity lemma (Theorem 3.1).

4.2 Our proposed IHCL++

As outlined in Chapter 3, Trevisan et al. showed that their regularity lemma (Theorem 3.1) implies Impagliazzo's original hardcore lemma (Theorem 4.9), and the results from [TTV09] predate the emergence of the field of algorithmic fairness. However, due to the equivalence between the multiaccuracy theorem that we described in Chapter 3, and given that we obtain our ++ theorems through the lenses of multicalibration, we will use the algorithmic fairness vocabulary when describing the theorems and proofs from [TTV09]. (Recall from Chapter 3 that we denote our stronger and more general theorems obtained through multicalibration with the symbol ++.)

In this chapter, we obtain a stronger and more general version of IHCL through a careful analysis of [TTV09]'s use of a multiaccurate predictor to prove IHCL. Their proof begins by invoking the multiaccurate theorem to obtain a multiaccurate predictor h. Then, they define the hardcore measure μ using this h, and then prove that μ satisfies the conclusion of IHCL by leveraging the multiaccuracy guarantees of h. We observe that a MA predictor h is *not* part of the IHCL statement —only $\mathcal{F}, g, \epsilon$, and $\mathcal{H}$ appear in the statement. Rather, h comes in as a tool in the proof, where the MA theorem is invoked. For this reason, when proposing our ICHL++ version, we decide to turn to multicalibrated *partitions*, rather than multicalibrated *predictors*. We introduced the definition of a multicalibrated partition in Section 2.3, where we also explained the relationship between an

MC predictor and an MC partition. We recall that the key idea behind this relationship is that the level sets of a predictor induce a partition of the domain $\mathcal{X}$. In this chapter (and subsequent chapters), we will also be making use of the notation introduced in Section 2.3 that formalizes the notion of a low-complexity partition.

For the ease of notation, we will be using the following abbreviations:

Definition 4.10. Given a set $P \subseteq \mathcal{X}$, any function $g\colon \mathcal{X} \to [0,1]$, and a distribution $\mathcal{D}$ on $\mathcal{X}$, we define the *balance* k_p of g on P to be

$$v_p := \mathop{\mathbb{E}}_{x \sim \mathcal{D}|_P}[g(x)], \qquad k_p := \min\{v_p, 1 - v_p\}.$$

By the definition of k_p, the parameter k_p is small precisely when the expected value of g is too close to 0 or too close to 1; i.e., when g is *imbalanced*. This is why we call k_p the "balance" of g on P. In particular, $k_p = 1/2$ corresponds g being perfectly balanced, while $k_p = 0$ corresponds to g being completely imbalanced (i.e., a constant function). The reason why we need to consider the parameter k_p in the proof of IHCL++. We will see that the indistinguishability guarantee within each set $P \in \mathcal{P}$ will degrade as ϵ/k_p. Therefore, the smaller k_p is, the worse the indistinguishability parameter becomes.

From Section 2.3, recall that we also need to consider the size parameter $\eta_p = |P|/|\mathcal{X}|$ of each $P \in \mathcal{P}$ (Definition 2.26). Because we are using the notion of approximate multicalibration, we will only be considering the sets $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$.

We can now introduce our IHCL++ statement:

Theorem 4.11 (IHCL++, measure version)**.** *Let $\mathcal{X}$ be a finite domain, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, let $g\colon \mathcal{X} \to \{0,1\}$ be an arbitrary function, and let $\epsilon, \gamma > 0$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ which satisfies that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, there exists a distribution $\mathcal{H}_P$ in P of density $2k_p$ in $\mathcal{U}_P$ such that g is $(\mathcal{F}, \epsilon/k_p)$-strongly hard on $\mathcal{H}_P$. That is,*

$$\forall f \in \mathcal{F}, \quad \Pr_{x \sim \mathcal{H}_P}[f(x) = g(x)] \leq \frac{1}{2} + \frac{\epsilon}{k_p}.$$

Interpretation of IHCL++. Before going into the proof of Theorem 4.11, we explain what our IHCL++ theorem entails and how it is a stronger and more general version of the original IHCL.

- In Theorem 4.11, we remove the δ-weakly hard assumption from the original IHCL theorem, but still obtain that g is strongly hard on some distribution. The caveat is that the lower bound on the density of each hardcore set $\mathcal{H}_p$ depends on the balance k_p of g on P. Namely, if g is an "uninteresting" function, then the density of the hardcore sets will be small. However, in our IHCL++, we can always guarantee strong hardness of g within each $P \in \mathcal{P}$ on $\mathcal{H}_P$.
- We provide a general lower bound for the density of the hardcore distribution $\mathcal{H}_P$ on each $P \in \mathcal{P}$ that depends on the expected value of g on P (i.e., on k_p). The parameter k_p is an abstraction of the original parameter δ in IHCL, given that in our Theorem 4.11 we have no assumption whatsoever on the function g, and hence we also have no δ parameter.

- In our ++ theorem, the original IHCL occurs both "locally" (on each $P \in \mathcal{P}$) and "globally" (on $\mathcal{X}$). Theorem 4.11 states that IHCL occurs locally; namely, we obtain a hardcore distribution $\mathcal{H}_P$ within each $P \in \mathcal{P}$. However, we can always "glue" the different hardcore measures together $\mathcal{H}_p$ in order to obtain a hardcore measure $\mathcal{H}$ on $\mathcal{X}$. Since g is strongly hard on each $\mathcal{H}_p$, g will also be strongly hard on the "glued" hardcore measure $\mathcal{H}$. In Section 4.2, we will show that if we glue the different hardcore measures $\mathcal{H}_p$ together weighted by their corresponding size parameter $\eta_p := |P|/|\mathcal{X}|$, and if we bring back the assumption that g is δ-weakly hard (which is the key assumption in the original IHCL statement), then the glued hardcore set $\mathcal{H}$ has density at least 2δ on $\mathcal{U}_{\mathcal{X}}$. That is, we have recovered the original IHCL statement from our IHCL++ theorem.

Throughout our proof of IHCL++, we will also summarize the [TTV09] proof of the original IHCL, highlighting the similarities and differences between the two proofs.

Proof of Theorem 4.11. The [TTV09] proof of IHCL begins by invoking the MA theorem to obtain a *multiaccurate* predictor h with respect to $\mathcal{F}, g$, and ϵ. In our case, we being by invoking the approximate-MC partition theorem (Theorem 2.29) with the same parameters ϵ, γ, and where $\mathcal{D}$ corresponds to the uniform distribution on $\mathcal{X}$. This gives us a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ satisfying

$$\left| \mathop{\mathbb{E}}_{x \sim P}[f(x) \cdot (g(x) - v_p)] \right| \leq \epsilon.$$

for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$. We remark that we can write $x \sim P$ instead of $\mathcal{P}(\mathcal{D})|_P$ because in this case $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$.

In [TTV09], the hardcore distribution $\mathcal{H}$ over the domain $\mathcal{X}$ is defined as

$$\mathcal{H}(x) := \frac{|g(x) - h(x)|}{\sum_{y \in \mathcal{X}} |g(y) - h(y)|}.$$

Trevisan et al. then show that 1) $\mathcal{H}$ is δ-dense in $\mathcal{U}_{\mathcal{X}}$, and that 2) g is strongly hard on $\mathcal{H}$. Inspired by their proof, we define the following probability distribution on each set $P \in \mathcal{P}$:

$$\mathcal{H}_P(x) := \frac{|g(x) - v_p|}{\sum_{y \in P} |g(y) - v_p|}.$$

We remark that, unlike in the multiaccuracy case of [TTV09], the denominator in the expression for $\mathcal{H}_P$ sums over the set P instead of over the entirety of the domain $\mathcal{X}$. As we now show as part of our density proof, the denominator of $\mathcal{H}_P$ is always non-zero, unless $v_p \in \{0, 1\}$. Recall that, by definition (Definition 4.10),

$$k_P = \min\{v_P, 1 - v_P\}.$$

Density guarantee. In order to prove that $\mathcal{H}_P$ has density $2k_P$ in P, by definition of density this corresponds to showing that

$$\mathcal{H}_p(x) \leq \frac{1}{2k_p \cdot |P|}.$$

In the case of [TTV09], they instead show that

$$\sum_{x \in X} |g(x) - h(x)| \geq \delta \cdot |X|,$$

which proves that their $\mathcal{H}$ is δ-dense in $\mathcal{U}_{\mathcal{X}}$. (Recall that [TTV09] does not recover the optimal δ density parameter.) Given that our k_P represents the density parameter of $\mathcal{H}_P$, it is helpful to draw the parallel between our parameter k_P and the δ parameter in [TTV09]. Their relationship will become clearer in when we recover the original IHCL from IHCL++ (Section 4.2, where we will see that $\mathbb{E}_{P \sim \mathcal{P}}[k_p] \geq \delta$ when we assume that g is δ-weakly hard.)

Let $G_0 = \{x \in \mathcal{X} \mid g(x) = 0\}$, and let $G_1 = \{x \in \mathcal{X} \mid g(x) = 1\}$. Then, it follows that

$$\begin{aligned}
\sum_{x \in P} |g(x) - v_p| &= \sum_{x \in P \cap G_1} |1 - v_p| + \sum_{x \in P \cap G_0} |0 - v_p| \\
&= \sum_{x \in P \cap G_1} (1 - v_p) + \sum_{x \in P \cap G_0} v_p = |P \cap G_1| \cdot (1 - v_p) + |P \cap G_0| \cdot v_p \\
&= v_p \cdot |P| \cdot (1 - v_p) + (1 - v_p) \cdot |P| \cdot v_p = 2v_p \cdot (1 - v_p) \cdot |P| \\
&= 2k_p \cdot (1 - k_p) \cdot |P|.
\end{aligned}$$

Moreover, $|g(x) - v_p| \leq 1 - k_p$. Therefore, we obtain that

$$\mathcal{H}_p(x) \leq \frac{1}{2k_p \cdot |P|},$$

which means that $\mathcal{H}_p$ has density $2k_p$, as required.

Therefore,

$$\sum_{x \in P} |g(x) - v_p| \geq k_p \cdot |P|,$$

as required.

Indistinguishability guarantee. We use the following identity, proven in [TTV09], which is applicable to any domain (i.e., they apply it to the domain $\mathcal{X}$, and we apply it on each $P \in \mathcal{P}$ instead):

$$|g(x) - h(x)| \cdot \mathbb{1}_{[f(x)=g(x)]} = \left[\left(f(x) - \frac{1}{2}\right) \cdot \Big(g(x) - h(x)\Big) + \frac{1}{2} \cdot |g(x) - h(x)|\right],$$

where $\mathbb{1}_{[f(x)=g(x)]}$ corresponds to the indicator random variable that returns 1 if and only if $f(x) = g(x)$ (and 0 otherwise).

In our case, $h(x) = v_p$ for all $x \in P$ for each $P \in \mathcal{P}$. Then by taking the expectation on both sides, and by applying the indistinguishability guarantee given by the ϵ-multicalibrated partition $\mathcal{P}$, we obtain that for each $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$,

$$\mathop{\mathbb{E}}_{x \sim P}[|g(x) - v_p| \cdot \mathbb{1}_{[f(x)=g(x)]}] \leq \epsilon + \frac{1}{2} \mathop{\mathbb{E}}_{x \sim P}[|g(x) - v_p|],$$

where the ϵ term follows from the assumption that $\mathcal{P}$ is an approximately multicalibrated partition, and thus $\left|\mathbb{E}_{x\sim P}[f(x)\cdot(g(x)-v_p)]\right| \le \epsilon$. Then, by the definition of $\mathcal{H}_P$, it follows that

$$\begin{aligned}\Pr_{x\sim\mathcal{H}_P}[f(x)=g(x)] &= \frac{\mathbb{E}_{x\sim P}[|g(x)-v_p|\cdot\mathbb{1}_{[f(x)=g(x)]}]}{\mathbb{E}_{x\sim P}[|g(x)-v_p|]} \le \frac{\epsilon+1/2\cdot\mathbb{E}_{x\sim P}[|g(x)-v_p|]}{\mathbb{E}_{x\sim P}[|g(x)-v_p|]}\\ &= \frac{1}{2}+\frac{\epsilon}{\mathbb{E}_{x\in P}[|g(x)-v_p|}.\end{aligned}$$

Since by the density guarantee we know that $\sum_{x\in P}|g(x)-v_p| \ge k_p\cdot|P|$, it follows that

$$\frac{1}{2}+\frac{\epsilon}{\mathbb{E}_{x\in P}[|g(x)-v_p|} \le \frac{1}{2}+\frac{\epsilon}{k_p} = \frac{1}{2}+\frac{\epsilon}{k_p}.$$

This concludes the proof of Theorem 4.11. □

Recovering the original IHCL from IHCL++

Having proved IHCL++, we now show how to recover the original IHCL theorem. The key idea is to "glue together" the hardcore sets $\mathcal{H}_p$ within each $P\in\mathcal{P}$, where in this gluing each $P\in\mathcal{P}$ is weighted according to its size parameter η_p of the set P.

Recall that in the IHCL statement, we assume that the function g is δ-weakly hard. Hence, we begin by showing that if g is δ-weakly hard, then "gluing" together the pieces P of the multicalibrated partition yields density δ.

Proposition 4.12. *Let $\mathcal{P}$ be a partition of $\mathcal{X}$ as in Theorem 4.11. Moreover, assume that g is δ-weakly hard with respect to $\mathcal{F}_{t,q}$ for some $\delta>0$, and suppose that $\eta_p\ge\gamma$ for all $P\in\mathcal{P}$. Then,*

$$\mathop{\mathbb{E}}_{P\sim\mathcal{P}}[k_p]\ge\delta.$$

We remark that we write $P\sim\mathcal{P}(\mathcal{D})$ as $P\sim\mathcal{P}$ because in this case $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$. We are selecting P with probability proportional to η_P.

Proof. We will argue by contradiction; hence assume that $\mathbb{E}_{P\sim\mathcal{P}}[k_p]<\delta$. We will show that this contradicts the fact that g is δ-weakly hard. More specifically, we show that we can construct an $f\in\mathcal{F}_{t,q}$ such that

$$\Pr[f(x)=g(x)]>1-\delta.$$

Let $f_m\in\mathcal{F}_{t,q}$, where $t=O(1/(\epsilon^4\gamma)\cdot\log(|\mathcal{X}|/\epsilon))$, $q=O(1/\epsilon^2)$, be the partition membership function for $\mathcal{P}$ as given by Definition 2.28. That is, $P_i=f_m^{-1}(i)$ for all of the k sets $P_i\in\mathcal{P}$. We define our f as follows:

- For each $x\in\mathcal{X}$, let $i=f(x)$.
- Let v_{p_i} be the expected value of the function g on P_i. If $v_{p_i}\le 1/2$, then return $f(x)=0$ for all $x\in P_i$. Otherwise, if $v_{p_i}>1/2$, then return $f(x)=1$.

(We always drop the subscript i when dealing with v_p and P, but in this case the subscript is needed to define this f.) Then, $f\colon\mathcal{X}\to\{0,1\}$. We claim that $f\in\mathcal{F}_{t,q}$ with the same parameters $t=$

$O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$. Indeed, let C_m be the oracle-aided circuit that computes f_m. It is enough that we hard-wire the values 0, 1 as described above. (To know whether it should be 0 or 1 for each $P_i \in \mathcal{P}$, we use a look-up table that contains the values v_{p_i}.) Hence, the circuit that computes f is of size $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon)) + |\mathcal{P}|$ and continues to have $q = O(1/\epsilon^2)$ oracle gates (the same as for f_m) [Bar22, §9.1.1.]. Since $|\mathcal{P}| = O(1/\epsilon)$ by Theorem 4.11, it follows that $f \in \mathcal{F}_{t,q}$, since the term $O(1/\epsilon)$ is absorbed into t.

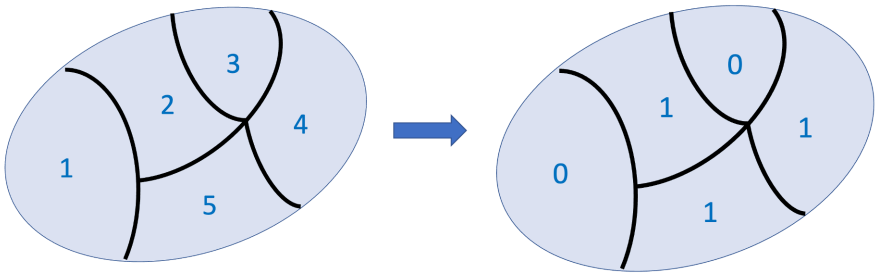

Figure 4.2: Illustration of the construction of the function f.

The intuitive meaning of the values 0 and 1 is the following: we want to show that f approximates g "quite well", in the sense that $\Pr[f(x) = g(x)] > 1 - \delta$. The above construction is saying that f is equal to 0 in all of the P_i such that $\mathbb{E}_{P_i}[g(x)] \leq 1/2$, and equal to 1 in all of the P_i such that $\mathbb{E}_{P_i}[g(x)] > 1/2$. We now show that this is indeed a good approximation of g; good enough that it contradicts the assumption that g is $(\mathcal{F}_{t,q}, \delta)$-weakly hard.

Fix some $P \in \mathcal{P}$, and as usual let $v_p := \mathbb{E}_{x \sim P}[g(x)]$. Since g is a boolean function and f equals the majority value of g on P by construction, it follows that

$$\Pr_{x \sim P}[f(x) = g(x)] = \max\{v_p, 1 - v_p\} = 1 - \min\{v_p, 1 - v_p\} = 1 - k_p,$$

since $f = \mathbf{0}$ when $v_p \leq 1/2$ and $f = \mathbf{1}$ when $v_p > 1/2$.

Because this holds for every $P \in \mathcal{P}$, when we consider the probability that $f(x) = g(x)$ over $\mathcal{X}$ it follows that

$$\Pr_{x \sim \mathcal{X}}[f(x) = g(x)] = 1 - \mathop{\mathbb{E}}_{P \sim \mathcal{P}}[k_p],$$

since

$$\Pr_{x \sim \mathcal{X}}[f(x) = g(x)] = \sum_P (1 - k_p) \cdot \frac{|P|}{|X|} = \mathop{\mathbb{E}}_{P \sim \mathcal{P}}[1 - k_p] = 1 - \mathop{\mathbb{E}}_{P \sim \mathcal{P}}[k_p].$$

Since by assumption $\mathbb{E}_{P \sim \mathcal{P}}[k_p] < \delta$, it follows that

$$\Pr_{x \sim \mathcal{X}}[f(x) = g(x)] \geq 1 - \delta,$$

which contradicts the $(\mathcal{F}_{t,q}, \delta)$-weakly hardness of g. □

In Proposition 4.12, we are assuming that $\eta_p \geq \gamma$ for all $P \in \mathcal{P}$ in order to make its proof cleaner. However, of course, we cannot be making this assumption. Instead, we should only be "gluing" together the pieces $P \in \mathcal{P}$ that have enough size and enough mass; i.e., such that η_p and k_p are larger than some threshold. In the case of the size parameter η_p, its threshold corresponds

to the γ parameter in the approximate MC definition. In the case of the balance parameter k_p, we introduce a new parameter τ:

Definition 4.13. Let $\gamma, \tau > 0$, and let $\mathcal{P}$ be a partition of the domain $\mathcal{X}$. We say that a set $P \in \mathcal{P}$ is (γ, τ)-*good* if $\eta_p \geq \gamma$ and $k_p \geq \tau$. We use the indicator random variable $\mathbb{1}_G$ to denote whether P is good. Namely, $\mathbb{1}_G(P)$ for $P \in \mathcal{P}$ returns 1 only if $\eta_p \geq \gamma$ and $k_p \geq \tau$; otherwise, it returns 0. (The letter G stands for "good".)

Given this definition, we now re-evaluate Proposition 4.12. Namely, the next fact follows directly from coupling the proof of Proposition 4.12 and Definition 4.13:

Corollary 4.14. *Let $\mathcal{P}, t, q, \epsilon$ as in Theorem 4.11, and let $\gamma, \tau > 0$. Moreover, assume that g is δ-weakly hard with respect to $\mathcal{F}_{t,q}$ for some $\delta > 0$. Then,*

$$\mathop{\mathbb{E}}_{P \sim \mathcal{P}}[k_p \cdot \mathbb{1}_G(P)] \geq \delta - \gamma \cdot |\mathcal{P}| - \tau = \delta - O(\gamma/\epsilon) - O(\tau),$$

where $\mathbb{1}_G(P)$ returns 1 only if $\eta_p \geq \gamma$ and $k_p \geq \tau$.

Recall that $|\mathcal{P}| = O(1/\epsilon)$ follows from Claim 2.11. We can now prove the original IHCL from IHCL++:

Proof of IHCL using IHCL++. Let $\mathcal{F}, \mathcal{X}, \epsilon, \delta$ be the assumption parameters in IHCL. We define the parameters $\epsilon' := \epsilon^2\delta$, $\gamma := \epsilon\epsilon'$, and invoke the IHCL++ theorem with these parameters ϵ', γ. By IHCL++, we obtain a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon'^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon'))$, $q = O(1/\epsilon'^2)$, $k = O(1/\epsilon')$ such that, for each $P \in \mathcal{P}$ where $\eta_p \geq \gamma = \epsilon\epsilon'$, there exists a distribution $\mathcal{H}_P$ in P of density $2k_p$ such that g is $(\mathcal{F}, \epsilon/k_p)$-hard on $\mathcal{H}_P$.

Let $\tau := \epsilon\delta$. By Corollary 4.14, when we only consider the $P \in \mathcal{P}$ that are (γ, τ)-good, we obtain that

$$\mathop{\mathbb{E}}_{P \sim \mathcal{P}}[k_p \cdot \mathbb{1}_G(P)] \geq \delta - O(\gamma/\epsilon') - \tau.$$

By plugging in the definition of each value $\epsilon' = \epsilon^2\delta$, $\gamma = \epsilon\epsilon'$, and $\tau = \epsilon\delta$, the expression $\delta - O(\gamma/\epsilon') - \tau$ simplifies to $\delta \cdot (1 - O(\epsilon))$.

We now construct a hardcore measure $\mathcal{H}$ on $\mathcal{X}$ as follows: we define $\mathcal{H}$ by "gluing up" the distributions $\mathcal{H}_p$ such that P is *good*. Formally, for each $x \in \mathcal{X}$,

$$\mathcal{H}(x) = \mathcal{H}_p(x) \cdot \mathbb{1}_G(P),$$

where P corresponds to the unique $P \in \mathcal{P}$ such that $x \in P$ (which is unique since $\mathcal{P}$ is a partition).

We now analyze (1) the density of $\mathcal{H}$, and (2) the hardness of g on $\mathcal{H}$. Since each $\mathcal{H}_p$ such that P is *good* has density $2k_p$ in $\mathcal{U}_P$ and $\mathbb{E}_{P \sim \mathcal{P}}[k_p \cdot \mathbb{1}_G(P)] \geq \delta \cdot (1 - O(\epsilon))$ by Proposition 4.12, it follows that $\mathcal{H}$ has density $2\delta \cdot (1 - O(\epsilon))$ in $\mathcal{U}_\mathcal{X}$.

For the hardness of g, we see that

$$\Pr_{x \sim \mathcal{H}}[f(x) = g(x)] = \mathop{\mathbb{E}}_{P \sim \mathcal{P}}\left[\Pr_{x \sim \mathcal{H}_p}[f(x) = g(x)] \cdot \mathbb{1}_G(P)\right] \leq \mathop{\mathbb{E}}_{P \sim \mathcal{P}}\left[\left(1/2 + \frac{\epsilon'}{k_p}\right) \cdot \mathbb{1}_G(P)\right] =$$

$$= \frac{1}{2} + \mathop{\mathbb{E}}_{P \sim \mathcal{P}}\left[\frac{\epsilon'}{k_p} \cdot \mathbb{1}_G(P)\right] \leq \frac{1}{2} + \frac{\epsilon'}{\tau}.$$

By plugging in the definitions of the parameters, namely $\epsilon' = \epsilon^2\delta$ and $\tau = \epsilon\delta$, we obtain that $\epsilon'/\tau = \epsilon$. Hence, we obtain that g is $(\mathcal{F}, \epsilon)$-strongly hard on $\mathcal{H}$.

Therefore, we have shown that $\mathcal{H}$ is a measure of density $2\delta \cdot (1 - O(\epsilon))$ in $\mathcal{U}_\mathcal{X}$ such that g is $(\mathcal{F}, O(\epsilon))$-strongly hard on $\mathcal{H}$. In order to recover the original IHCL statement, we observe that we can modify $\mathcal{H}$ in order to make its density at least 2δ in $\mathcal{U}_\mathcal{X}$ while maintaining the $(\mathcal{F}, O(\epsilon))$-strong hardness of g on $\mathcal{H}$. Namely, we can arbitrarily increase the value $\mathcal{H}(x)$ for some x in order to get to density 2δ. However, this modification can only change $\Pr_{x\sim\mathcal{H}}[f(x) = g(x)]$ by at most $O(\epsilon)$. Therefore, despite this modification, g is $(\mathcal{F}, O(\epsilon))$-strongly hard on $\mathcal{H}$. □

4.3 Sets and measures

There are two possible ways of describing the hardcore set of inputs, as it was already described in [Imp95; Hol05]. One is to find a hardcore *measure*, which corresponds to the distribution μ in Theorem 4.9. Notice that μ is defined over all the domain $\mathcal{X}$: each point $x \in \mathcal{X}$ receives a probability mass $\mu(x)$. That is, Theorem 4.9 corresponds to the measure version of Impagliazzo's Hardcore Lemma.

On the other hand, we can compute a hardcore *set*, which corresponds to a subset of the inputs $\mathcal{X}$. To obtain a hardcore set H from a hardcore measure μ, we construct H probabilistically as follows: for each $x \in \mathcal{X}$, we add x to H with probability $\mu(x)$. Then, the fact that μ is δ-dense with respect to $U_\mathcal{X}$ implies that H is δ-dense in $\mathcal{X}$ on expectation (i.e., that $|H| \leq \delta|\mathcal{X}|$). To show this, we proceed via a probabilistic method argument and a Chernoff bound. We formalize this idea in Lemma 4.16. For details, see Section 6 in [Imp95], Section 2.1.1 in [Hol05], or Section 4.4 in [KS03].

For our proofs in this chapter, we will proceed with the measure version of the theorem. However, we will also state the theorems in their set versions: while the measure version is better for some applications, the set version is easier to visualize and is generally more intuitive. We recall that $d(\mathcal{H})$ denotes the density of a measure $\mathcal{H}$.

The idea behind the conversion between sets and measures is originally due to Impagliazzo, but we follow the proof in [KS03, §4.4]. For that, we need to clarify the meaning of a hardcore *set* (as opposed to a hardcore *distribution*). Recall that every $S \subseteq \mathcal{X}$ has a corresponding measure given by the associated characteristic function $\chi_S \colon \mathcal{X} \to \{0, 1\}$. That is, $\chi_S(x) = 1$ if $x \in S$, and $\chi_S(x) = 0$ if $x \notin S$. Given that every measure induces a probability distribution (Definition 4.4), S induces a probability distribution $\mathcal{D}_S$, defined as

$$\mathcal{D}_S(x) = \frac{\chi_S(x)}{\sum_{x\in\mathcal{X}} \chi_S(x)}.$$

The natural definition of a hardcore set is then the following:

Definition 4.15 (Hardcore set)**.** Given a class $\mathcal{F}$ of functions $f \colon \mathcal{X} \to \{0, 1\}$, an arbitrary function $g \colon \mathcal{X} \to \{0, 1\}$, and $\epsilon > 0$, we say that g is *ϵ-strongly hard* with respect to $\mathcal{F}$ on a set $H \subset \mathcal{X}$ if, for all $f \in \mathcal{F}$,

$$\Pr_{x\sim H}[f(x) = g(x)] \leq \frac{1}{2} + \epsilon.$$

In that case, we say that H is an *$(\mathcal{F}, \epsilon)$-hardcore set* for g.

That is, g is very hard to approximate inside inside of the set $H \subseteq \mathcal{X}$.

We can now state the formal conversion from a hardcore measure to a hardcore set. The reader should feel free to skip the proof, as it is not essential for the rest of this chapter.

Lemma 4.16 (From measures to sets [KS03, Lemma 24])**.** *Let $\mathcal{X} = \{0,1\}^n$, $g\colon \mathcal{X} \to \{0,1\}$ an arbitrary function, $\epsilon, \delta > 0$, such that $\delta \geq 1/|\mathcal{X}|^{1/2}$, and $\mathcal{F}$ a class of functions $f\colon X \to \{0,1\}$ such that $\log(|\mathcal{F}|) \leq |\mathcal{X}|\epsilon^2\delta^2$. Suppose that μ is an $(\mathcal{F},\epsilon)$-hardcore measure for g such that $d(\mu) \geq \delta$. Then there exists a set H with $|H| \geq \delta \cdot |\mathcal{X}|$ such that H is an $(\mathcal{F}, 4\epsilon)$-hardcore set for g.*

Proof. Given the measure μ, we construct the claimed set H as follows: for each $x \in \mathcal{X} = \{0,1\}^n$, include x to H with probability $\mu(x)$. This is a non-constructive set H, but we are still able to argue about its density. We denote the characteristic function of H by χ_H.

Let $f \in \mathcal{F}$ and let $t(x)$ be an arbitrary function. We now use the following fact, where without loss of generality we view functions f and t as taking values in $\{-1,1\}$:

Lemma 4.17. *Let $\rho > 0$. Then,*

$$\Pr_{\mathcal{D}_\mu}[f(x) = t(x)] = \frac{1}{2} + \rho \iff \sum_{x\in\{0,1\}^n} \mu(x)f(x)t(x) = 2\rho|\mu|,$$

where $\mathcal{D}_\mu$ denotes the probability distribution induced by μ and $|\mu| = \sum_{x\in\mathcal{X}} \mu(x)$.

This is a direct consequence from the definition of $|\mu|$. Next, by the definition of χ_H, it follows that

$$\mathop{\mathbb{E}}_{x\sim H}[\chi_H(x)] = \mu(x).$$

By linearity of expectation, it follows that

$$\mathop{\mathbb{E}}_{x\sim H}\Big[\sum_{x\in\mathcal{X}} \chi_H(x)f(x)g(x)\Big] = \sum_{x\in\mathcal{X}} \mu(x)f(x)g(x).$$

Then, by applying Lemma 4.17 with $t := g$, it follows that

$$\sum_{x\in\mathcal{X}} \mu(x)f(x)g(x) \leq 2\epsilon|\mu|,$$

since by assumption μ is an $(\mathcal{F}_t,\epsilon)$-hardcore measure for g. By definition, $\Pr_{x\sim D_\mu}[f(x) = g(x)] \leq \epsilon$, and hence the parameter ρ in Lemma 4.17 corresponds to ϵ.

Next, we use Hoeffding's inequality:

Claim 4.18 (Hoeffding's tail bound)**.** *Let $X_1, \ldots, X_N$ be independent random variables with $X_i \in [a,b]$ for all i. Then, for all $t \geq 0$,*

$$\Pr\Big[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\Big] \leq \exp\Big(\frac{-2nt^2}{(b-a)^2}\Big).$$

In our case, for each value of $x \in \mathcal{X}$ the quantity $X_x := \chi_H(x)f(x)g(x)$ is a random variable in the interval $[-1,1]$. Hence, each of these random variables corresponds to one X_i in Hoeffding's

bound. We just showed that

$$\mathop{\mathbb{E}}_{x\sim\mathcal{X}}\Big[\sum_x \chi_H(x)f(x)g(x)\Big] \le 2\epsilon|\mu| = 2\epsilon\cdot|\mathcal{X}|\cdot d(\mu),$$

since by definition of density of a measure (Definition 4.5) we know that

$$d(\mu) = \frac{|\mu|}{|\mathcal{X}|}.$$

Set $t := 2\epsilon d(\mu)$. Since $X_x \in [-1,1]$ for all $x \in \mathcal{X}$, it follows that $(b-a)^2 = 4$. Then, by Hoeffding's tail bound it follows that

$$\Pr\Big[\frac{1}{|\mathcal{X}|}\sum_x X_x \ge 4\epsilon d(\mu)\Big] \le \exp\Big(\frac{-2\cdot|\mathcal{X}|(2\epsilon d(\mu))^2}{4}\Big).$$

Since by assumption $d(\mu) \ge \delta$, it follows that the above probability is less than $\exp(-2\cdot|\mathcal{X}|\cdot\epsilon^2\delta^2)$. Since by assumption $|\mathcal{F}| \le 2^{|\mathcal{X}|\cdot\epsilon^2\cdot\delta^2} \ll \frac{1}{10}\exp(2\cdot|\mathcal{X}|\cdot\epsilon^2\delta^2)$, and hence by the the union bound this implies that the probability that there exists some $f\in\mathcal{F}$ such that $\sum_x \chi_H(x)f(x)g(x) \ge 4\epsilon|\mu|$ is less than $1/10$.

Next, applying the Hoeffding bound to $|H|$ (which is a sum of $|\mathcal{X}|$ independent random variables), by similar calculations and using the assumptions that $d(\mu)\ge\delta$ and $\delta \ge 1/|\mathcal{X}|^{1/2}$, it follows that $|H| \ge 2\delta\cdot|\mathcal{X}|$.

Lastly, putting everything together, we conclude that there exists some set H such that

(1) $|H| \ge \delta\cdot|\mathcal{X}|$,

(2) $\sum_x \chi_H(x)f(x)g(x) \le 4\epsilon|\mu| \le 8\delta|H| = 8\epsilon|\chi_H|$.

Using Lemma 4.17 with $t := g$ and $\rho := 4\epsilon$ we thus obtain that

$$\Pr_{x\sim H}[f(x) = g(x)] \le \frac{1}{2} + 4\epsilon$$

for all $f\in\mathcal{F}$. Hence, H is an $(\mathcal{F}, 4\epsilon)$-hardcore set for g, as we wanted to show. □

Remark 4.19. In Lemma 4.16, we used the assumption $\log(|\mathcal{F}|) \le |\mathcal{X}|\epsilon^2\delta^2$. We remark that in the case where $\mathcal{F}$ corresponds to circuits of size s and $\mathcal{X} = \{0,1\}^n$, this assumption is always true. Namely, in that case, $|\mathcal{F}| \le 2^{2ns}$ always holds. This is because, by a counting argument, in order to specify a circuit it suffices to specify, for each of the s gates of the circuit, the two inputs and the label of the gate.

Set version of ICHL++

Given that we can transform measures into sets via a probabilistic method argument (as described in Section 4.3), we can now state the corresponding versions of our IHCL++ theorem (measure version; Theorem 4.11) following the same transformation. While these hardcore sets are non-constructive due to the nature of the probabilistic method argument used in Section 4.3, it is more

intuitive to grasp the IHCL++ by visualizing hardcore sets rather than hardcore measures (see, e.g., Figure 4.3). Moreover, we use the set version of IHCL and IHCL++ in Chapter 5.

Theorem 4.20 (IHCL++, Set version)**.** *Let $\mathcal{X}$ be a finite domain, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, let $g\colon \mathcal{X} \to \{0,1\}$ be an arbitrary function, let $\epsilon, \gamma > 0$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ which satisfies that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$ and such that $|\mathcal{F}| \leq \frac{1}{10}\exp(2|\mathcal{X}|\epsilon^2 k_p^2)$, there exists a set $H_p \subseteq P$ of density $|H_p|/|P| \geq 2k_p$ such that H_p is an $(\mathcal{F}, \epsilon/k_p)$-hardcore set for g. That is,*

$$\forall f \in \mathcal{F}, \quad \Pr_{x \in H_p}[f(x) = g(x)] \leq 1/2 + \frac{\epsilon}{k_p}.$$

Proof. We use the ICHL++ measure version theorem (Theorem 4.11) with $\epsilon/4$ to obtain a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ such that there exists a distribution $\mathcal{H}_p$ of density $2k_p$ in $\mathcal{U}_P$ for each $P \in \mathcal{P}$ such that g is $(\mathcal{F}, \epsilon/(\eta_p \cdot k_p))$-hard on $\mathcal{H}_P$. Then, we apply Lemma 4.16 to each of the $\mathcal{H}_p$ to obtain the corresponding set $H_p \subseteq P$. □

Visual interpretation of IHCL++. Having translated the IHCL++ to hardcore sets rather than hardcore measures, we can now more easily visualize and interpret the IHCL++ theorem and corollary. In the original IHCL statement, we assume that g is δ-weakly hard and obtain a hardcore set H for g that occupies at least a 2δ-fraction of the space $\mathcal{X}$. In our IHCL++, we do *not* assume that g is δ-weakly hard, and we obtain many "little" hardcore sets H_p, one per set P in the partition $\mathcal{P}$, such that each of these sets occupies at least a $2k_p$-fraction of the space $|P|$ (i.e., of the set P to which they belong). The k_p parameter corresponds to the balance of g in the set P: the closer the expected value of g on P is to 0 or 1, the smaller the k_p parameter is, and hence the "less interesting" the hardcore set is (in the sense that it might occupy a very small fraction of the space).

However, when we bring back the assumption that g is δ-weakly hard, we proved that then the average value of the k_p parameter (over all $P \in \mathcal{P}$) is at least δ (Propostion 4.12). Then, by "gluing" all of the "little" hardcore sets together, we obtain a large hardcore set – large in the sense that it occupies at least a 2δ fraction of the domain $|\mathcal{X}|$, as in the original IHCL theorem. However, unlike the original IHCL theorem, each P continues to have its own "little" hardcore set.

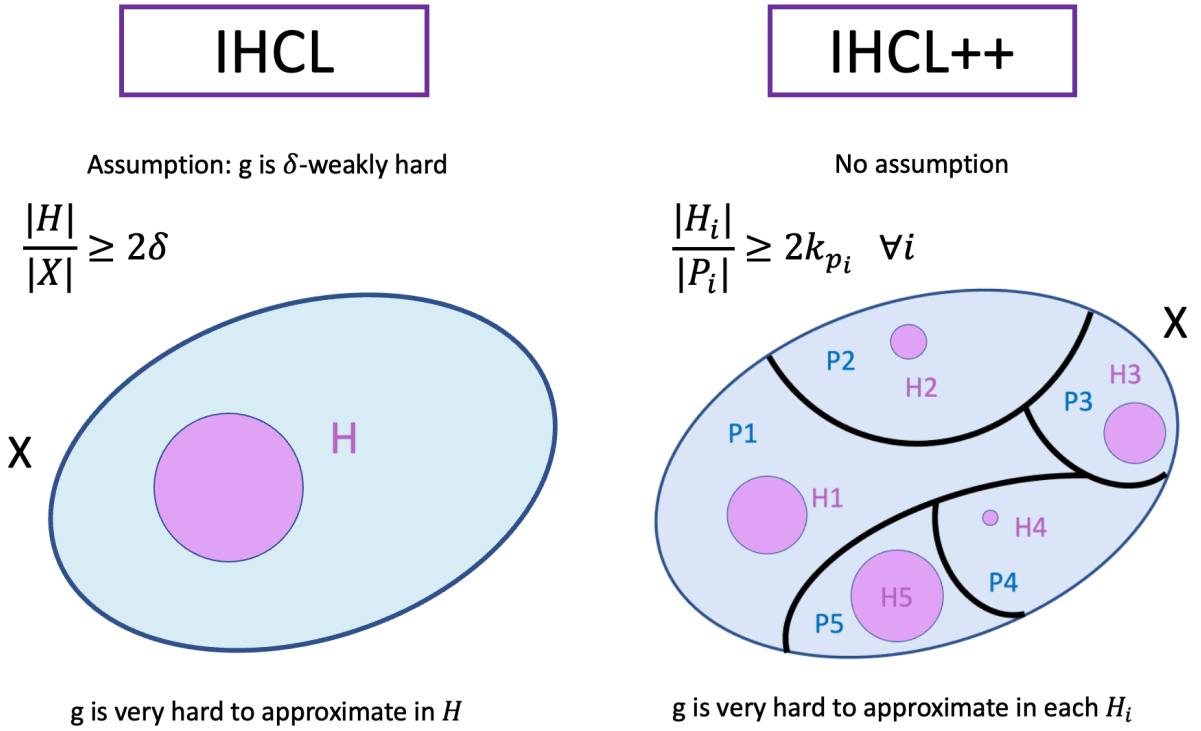

Figure 4.3: Visual representation of the difference between the original IHCL and our ICHL++.

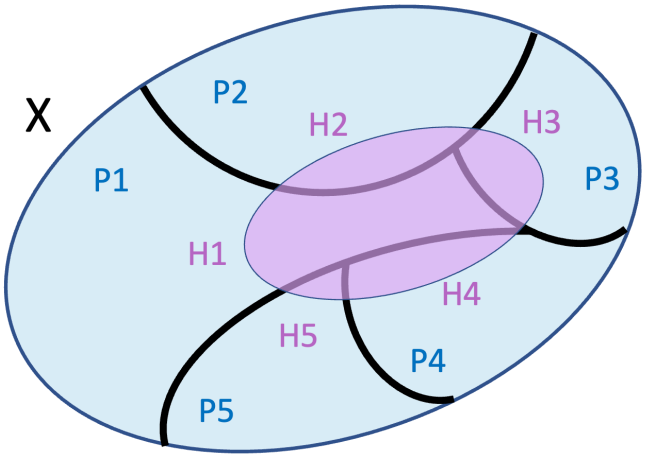

Figure 4.4: Visual representation of the recovery of the original IHLC from IHCL++. The density of H_i within P_i might still be "uninteresting", but when we glue all of the H_i together, we obtain a hardcore set $H \subseteq \mathcal{X}$ that occupies at least a 2δ fraction of the space.

4.4 Multicalibration gives us indistinguishability "for free"

In this section, we further explore the distinctions between the original IHCL statement and our ICHL++ through the lenses of the differences between MA and MC. A natural question that arises when one encounters the IHCL++ statement is: How is it possible that the IHCL++ theorem is able to obtain hardcore sets for an *arbitrary* function g rather than for a δ-weakly hard function g? Returning to our explanation at the beginning of this chapter, we should think of a hardcore set

as subset of the domain $\mathcal{X}$ where $\mathcal{X}$ is very hard to approximate by functions $f \in \mathcal{F}$. In the hardcore set, the function g is behaving "like a random function", because the distinguishers $f \in \mathcal{F}$ cannot guess the outputs of $g(x)$ with probability larger than $1/2 + \epsilon$. We make this idea precise in Claim 4.23.

In order to provide intuition for why IHCL++ works for an arbitrary function, we have to return to the notion of multicalibration. The key to our proof of IHCL++ is to use a multicalibrated partition $\mathcal{P}$, which, as we explained in Chapter 2 (more concretely, in Claim 2.27) is induced by a multicalibrated predictor. Let us recall the definition of an $(\mathcal{F}, \epsilon)$-multicalibrated partition $\mathcal{P}$: each $P \in \mathcal{P}$ satisfies

$$\left| \mathop{\mathbb{E}}_{x \sim P}[f(x) \cdot (g(x) - v_p)] \right| \leq \epsilon,$$

where $v_p := \mathbb{E}_P[g(x)]$ and $\eta_p := |P|/|\mathcal{X}|$. (For simplicity, assume that $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$.)

Meanwhile, if h is an $(\mathcal{F}, \epsilon)$-multiaccurate predictor for g, this implies that

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{X}}[f(x) \cdot (g(x) - h(x))] \right| \leq \epsilon.$$

The key to understanding the "power" of a multicalibrated partition $\mathcal{P}$ is to realize that MC is not just "MA on each level set". Not only did we swap $x \sim \mathcal{X}$ for $x \sim P$ below the expectation, another difference is that, in the case of MC, v_p is a *constant*, whereas the $h(x)$ in MA is *not*. Crucially, this implies that, within each level set P, the function g is ϵ-indistinguishable from a Bernoulli random variable of parameter v_p. Formally:

Lemma 4.21 (g is indistinguishable from a Bernoulli within each level set)**.** *Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a family of functions $f \colon \mathcal{X} \to \{0,1\}$, $g \colon \mathcal{X} \to \{0,1\}$ an arbitrary function, and $\epsilon > 0$. Let $\mathcal{P}$ be an $(\mathcal{F}, \epsilon)$-multicalibrated partition for g. Then, within each $P \in \mathcal{P}$, g is $(\mathcal{F}, \epsilon)$-indistinguishable from the random variable $X_p \sim \mathrm{Bern}(v_p)$.*

We recall the definition of $(\mathcal{F}, \epsilon)$-indistinguishability from Chapter 2:

Definition 4.22. Given $\mathcal{X}, \mathcal{F}, \epsilon > 0$, a distribution $\mathcal{D}$ on $\mathcal{X}$, and two arbitrary functions $g, h \colon \mathcal{X} \to [0,1]$, we say that g and h are $(\mathcal{F}, \epsilon)$-*indistinguishable* if

$$\left| \mathop{\mathbb{E}}_{x \sim \mathcal{X}}[f(x) \cdot (g(x) - h(x))] \right| \leq \epsilon.$$

We also remark that $X \sim \mathrm{Bern}(v_p)$ means that X is distributed as a Bernoulli random variable of parameter v_p. This means that $\Pr[X = 1] = v_p$ and $\Pr[X = 0] = 1 - v_p$. It is useful to think of $X \sim \mathrm{Bern}(v_p)$ as a coin flip with a biased coin, where the bias corresponds to v_p.

Proof of Lemma 4.21. Since $\mathcal{P}$ is an $(\mathcal{F}, \epsilon)$-MC partition for g, it follows that

$$\left| \mathop{\mathbb{E}}_{x \sim P}[f(x) \cdot (g(x) - v_p)] \right| \leq \epsilon,$$

for all $f \in \mathcal{F}$ and $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$. Then, the claim follows directly from the fact that $X_p \sim \mathrm{Bern}(v_p)$ implies that $\mathbb{E}[X_p] = v_p$. □

This is the key property of a multicalibrated partition: we have managed to partition $\mathcal{X}$ into disjoint sets P such that within each P, g is indistinguishable from a Bernoulli of parameter v_P. In a way, we can think of this as defining g as piece-wise Bernoulli's, which are extremely simple functions: being indistinguishable from a Bernoulli is the computational analogue of being equal to a constant. Hence, we can think of these $|\mathcal{P}|$ Bernoulli random variables as the "scaffoldings" of g.

Having established that an MC partition of $\mathcal{X}$ corresponds to a decomposing $\mathcal{X}$ into parts $P \in \mathcal{P}$ such that g is indistinguishable from a Bernoulli random variable in each P, we now explain why we should expect this to imply that g is strongly hard in some region H_p within each P.

To do so, we begin by analyzing the case where $X \sim \text{Bern}(1/2)$. I.e., this would correspond to a set $P \in \mathcal{P}$ where $v_p := \mathbb{E}_P[g(x)] = 1/2$. We are interested in this case because a random function corresponds to a Bernoulli random variable of parameter $1/2$. Again, we think of $\text{Bern}(v_p)$ as a coin flip with bias v_p. Hence, in the case where $v_p = 1/2$, this corresponds to an unbiased coin flip, and hence the coin is behaving "randomly". Intuitively, this is why a random function yields strong hardness: if we were trying to guess the outputs of a boolean random function / a Bernoulli of parameter $1/2$, we would expect to be correct on half of the cases on average. Formally:

Claim 4.23. *Let $\mathcal{X}$ be a finite domain, $\mathcal{F}$ a class of distinguishers $f\colon \mathcal{X} \to \{0,1\}$, $g\colon \mathcal{X} \to [0,1]$, and $\epsilon > 0$. If g is $(\mathcal{F}, \epsilon)$-indistinguishable from the constant $1/2$ function, then g is $(\mathcal{F}, \epsilon)$-strongly hard.*

Proof. This follows from the usual equivalence between pseudorandomness and unpredictability. Namely:

$$\Pr_{x\sim\mathcal{X}}[f(x) = g(x)] = 2 \mathop{\mathbb{E}}_{x\sim\mathcal{X}}[(f(x) - 1/2)(g(x) - 1/2)] + 1/2.$$

Note that this is the same equivalence that we used in the indistinguishability part of the proof to Theorem 4.11. □

This is why an MC partition is giving us "indistinguishability for free": If $v_p = 1/2$ in some $P \in \mathcal{P}$, then g is already $(\mathcal{F}, \epsilon)$-strongly hard over P. If $v_p \neq 1/2$, then obtaining strong hardness is not as direct (as we study in the following section), but one would expect that we can get it from moving v_p closer to $1/2$. Still, the "real work" comes from the multicalibrated partition, which ensures that g is indistinguishable to a Bernoulli random variable on each $P \in \mathcal{P}$. To obtain strong hardness, it is a matter of "shifting" the v_p parameter to $1/2$.

In the following section, we use these ideas and observations about a multicalibration partition to provide an alternative proof to IHCL++. Importantly, this proof does recover the 2δ optimal density paramter, which we did *not* achieve in our first proof for IHCL++ (Section 4.2).

4.5 An alternative proof to IHCL++

By using the ideas described in Section 4.4, namely, how MC provides "indistinguishability for free, we can obtain a new and different proof for our IHCL++ theorem (Theorem 4.11). We believe that presenting this proof is valuable because it uses a completely different technique than our first proof. (Recall that our first proof followed the structure of the proof by Trevisan et al. that a multiaccurate predictor implies IHCL.)

We begin by observing the following: Consider the set version of our IHCL++. In the case where $v_p = 1/2$ for some $P \in \mathcal{P}$, this implies that there exists a hardcore set H_p in P of density

$|H_p|/|P| \geq 2k_p = 2 \cdot \min\{1/2, 1/2\} = 1$. This means that the hardcore set is *the entire set* P. We first see that this does indeed make sense; that is, that the entire level set is a hardcore set. Intuitively, this is explained by the ideas that we laid out in Section 4.4. Namely, when $v_p = 1/2$, the multicalibration condition implies that g is $(\mathcal{F}, \epsilon)$-indistinguishable from the constant $1/2$ function. By Claim 4.23, this implies that g is already $(\mathcal{F}, \epsilon)$-strongly hard on the entire set, and therefore it makes sense that $H_p = P$ in this case.

We now formally prove that this is indeed the case.

1. Changing the ranges of f and g. Assume that $\eta_p \geq \gamma$ and $v_p = 1/2$. Let $f, g \colon \mathcal{X} \to \{0, 1\}$, and let $\tilde{f}, \tilde{g} \colon \mathcal{X} \to \{-1, 1\}$. We change between f, g and $\tilde{f}, \tilde{g}$ as follows:

$$\tilde{f}(x) := 2f(x) - 1, \qquad f(x) := \frac{1}{2}\tilde{f}(x) + \frac{1}{2}.$$

We show that if $v_p = 1/2$ on a set $P \in \mathcal{P}$, then g is $(\mathcal{F}, \eta_p\epsilon)$-strongly hard on P. By the MC condition on set $P \in \mathcal{P}$, we know that

$$\left| \mathop{\mathbb{E}}_{P}[f(x) \cdot (g(x) - 1/2)] \right| \leq \epsilon,$$

since $v_p = 1/2$.

Here, all expectations and probabilities are taken with respect to the uniform distribution on P, i.e., each point has probability mass $1/|P|$. By the transformations between $f, \tilde{f}, g, \tilde{g}$ above, it follows that

$$g(x) - \frac{1}{2} = \frac{1}{2} \cdot \tilde{g}(x),$$
$$f(x) = \frac{1}{2}\tilde{f}(x) + \frac{1}{2}.$$

Plugging these into the MC condition on P, the expression becomes

$$\left| \mathop{\mathbb{E}}_{P}\left[\left(\frac{1}{2} \cdot \tilde{f}(x) + \frac{1}{2} \right) \cdot \left(\frac{1}{2} \cdot \tilde{g}(x) \right) \right] \right| \leq \epsilon.$$

In order to simplify the expression, since $\mathcal{F}$ is arbitrary, we instead apply the MC condition to the distinguisher $2f(x) - 1$. Then, the MC condition with $\tilde{f}, \tilde{g}$ becomes

$$\left| \mathop{\mathbb{E}}_{P}\left[\left(2 \cdot \left(\frac{1}{2} \cdot \tilde{f}(x) \right) \right) \cdot \left(\frac{1}{2} \cdot \tilde{g}(x) \right) \right] \right| \leq \epsilon,$$

which simplifies to

$$\left| \mathop{\mathbb{E}}_{P}[\tilde{f}(x) \cdot \tilde{g}(x)] \right| \leq \epsilon. \tag{4.24}$$

Next, we show that g is $(\mathcal{F}, 1/2 - \eta_p\epsilon)$-hard on P. For the sake of contradiction, suppose that

$$\Pr_{x \sim P}[f(x) = g(x)] > \frac{1}{2} + \epsilon.$$

Then,

$$\left|\mathop{\mathbb{E}}_{P}[\tilde{f}(x)\cdot\tilde{g}(x)]\right| = 2\mathop{\Pr}_{x\sim P}[f(x)=g(x)]-1>2\epsilon,$$

which contradicts Equation 4.24. Therefore, g is $(\mathcal{F},\epsilon)$-strongly hard in the entire level set P when its expected value is $1/2$, as we wanted to show.

2. Balancing g using μ. We now proceed to the general case, i.e., when $\mathbb{E}_P[g(x)]=v_P$ for any $v_P\in[0,1]$ (rather than the restricted case $v_p=1/2$. Our idea is to use the measure μ_p to "balance" g so that its expected value over μ_p is $1/2$, and then we can apply the previous reasoning to show that g is $(\mathcal{F},\epsilon/v_p(1-v_p))$-strongly hard.

Let $P\in\mathcal{P}$. We want to define a distribution μ_p over P such that

$$\mathop{\mathbb{E}}_{x\sim\mu_p}[g(x)]=1/2.$$

We claim that we can do so by defining μ_p as follows:

$$\mu_p(x)=\begin{cases}\dfrac{1}{2v_p}\cdot\dfrac{1}{|P|} & \text{if } g(x)=1,\\[2ex] \dfrac{1}{2(1-v_p)}\cdot\dfrac{1}{|P|} & \text{if } g(x)=0.\end{cases}$$

First, we check that $\mu_p(x)$ is indeed a probability distribution. Let $G_0=\{x\in\mathcal{X}\mid g(x)=0\}$ and $G_1=\{x\in\mathcal{X}\mid g(x)=1\}$.

$$\begin{aligned}\sum_{x\in P}\mu_p(x)&=\sum_{x\in\cap G_1}\mu_p(x)+\sum_{x\in P\cap G_0}\mu_P(x)=\sum_{x\in\cap G_1}\frac{1}{2v_p}\cdot\frac{1}{|P|}+\sum_{x\in P\cap G_0}\frac{1}{2(1-v_p)}\cdot\frac{1}{|P|}\\&=v_p\cdot|P|\cdot\frac{1}{2v_p}\cdot\frac{1}{|P|}+(1-v_P)\cdot|P|\cdot\frac{1}{2(1-v_p)}\cdot\frac{1}{|P|}=\frac{1}{2}+\frac{1}{2}=1.\end{aligned}$$

Moreover, it is clear that $\mu_p(x)\in[0,1]$ for all $x\in P$. Hence, μ_p is indeed a probability distribution.

Next, we show that $\mathbb{E}_{x\sim\mu_p}[g(x)]=1/2$.

$$\mathop{\mathbb{E}}_{x\sim\mu_p}[g(x)]=\sum_{x\in P}\mu_p(x)\cdot g(x)=\sum_{x\in P\cap G_1}\mu_p(x)=\sum_{x\in P\cap G_1}\frac{1}{2v_p}\cdot\frac{1}{|P|}=v_p\cdot|P|\cdot\frac{1}{2v_p}\cdot\frac{1}{|P|}=\frac{1}{2}.$$

Finally, we show that μ_p has the required density. Let $\overline{\mu}_p$ be the non-normalized version of μ; i.e., $\overline{\mu}_p(x)=1/(2v_p)$ if $g(x)=1$ and $\overline{\mu}_p(x)=1/(2(1-v_p))$ if $g(x)=0$. Then, we want to show that $\sum_{x\in P}\overline{\mu}(x)\geq 2k_p\cdot|P|$. By a similar calculation:

$$\sum_{x\in P}\overline{\mu}(x)=\sum_{x\in P\cap G^1}\overline{\mu}(x)+\sum_{x\in P\cap G^0}\overline{\mu}(x)=\frac{1}{2v_p}\cdot v_p\cdot|P|+\frac{1}{2(1-v_p)}\cdot(1-v_p)\cdot|P|=|P|.$$

Then, $|P|\geq 2k_p\cdot|P|$ holds if and only if $\frac{1}{2}\geq k_p$, which is always true since $k_p=\min\{v_p,1-v_p\}$ by definition, and $v_p\in[0,1]$.

Therefore, we now know that

- $\mathbb{E}_{x \in P}[g(x)] = v_p$.
- $\mathbb{E}_{x \sim \mu_P}[g(x)] = 1/2$.

By the approximate MC condition, we know that, for each $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$,

$$\left| \mathop{\mathbb{E}}_{P}[f(x) \cdot (g(x) - v_p)] \right| \leq \epsilon.$$

We now consider the MC inequality by splitting it between the points in $\mathcal{X}$ where g is 1 and the points in $\mathcal{X}$ where g is 0. That is:

$$\begin{aligned}\left| \mathop{\mathbb{E}}_{P}[f(x) \cdot (g(x) - v_p)] \right| &= v_p \cdot \mathop{\mathbb{E}}_{G_1 \cap P}[f(x) \cdot (1 - v_p)] + (1 - v_p) \cdot \mathop{\mathbb{E}}_{G_0}[f(x) \cdot v_p] \\ &= v_p \cdot (1 - v_p) \cdot \left| \mathop{\mathbb{E}}_{G_1 \cap P}[f(x)] - \mathop{\mathbb{E}}_{G_0 \cap P}[f(x)] \right| \leq v_p \cdot (1 - v_p) \cdot \epsilon,\end{aligned}$$

by the definition of μ. Intuitively, μ was defined precisely so that the distinguishers cannot tell whether we are sampling from $G_1 \cap P$ or from $G_0 \cap P$.

Therefore, we achieve indistinguishability with respect to the following ϵ':

$$\epsilon' := \frac{\epsilon}{v_p(1 - v_p)}.$$

Let μ_0 correspond to the restriction of μ_p on the domain $\{x \in P \mid g(x) = 0\}$, and let μ_1 correspond to the restriction of μ_p on the domain $\{x \in P \mid g(x) = 1\}$. Then, it follows that

$$\begin{aligned}\Pr_{\mu}[f(x) = g(x)] &= \frac{1}{2} \Pr_{\mu_1}[f(x) = 1] + \frac{1}{2} \Pr_{x \in \mu_0}[f(x) = 0] \\ &= \frac{1}{2} + \mathop{\mathbb{E}}_{\mu_1}[f(x)] - \mathop{\mathbb{E}}_{\mu_0}[f(x)] = \frac{1}{2} + \epsilon',\end{aligned}$$

as we wanted to show.

5

Characterizations of Pseudoentropy

Computational analogues of information-theoretic notions have given rise to some of the most interesting phenomena in complexity and cryptography. For example, computational indistinguishability, which is the computational analogue of statistical distance, enabled bypassing Shannon's impossibility results on perfectly secure encryption, and provided the basis for the computational theory of pseudorandomness.

Vadhan & Zheng [VZ13]

Information theory studies the quantification and communication of *information*. Claude Shannon was one of the main founders of the field in the 1940s, and it has since then become a key element in many applications, such as data compression or error correcting codes. A fundamental notion in information theory is that of *entropy* of a random variable X, which can be thought of as the amount of randomness in X.

In computational complexity and cryptography, a key development has been the study of *computational analogues* of concepts from information theory. Before we use any concepts from information theory, we have already encountered this phenomenon in Chapter 4: namely, we explained how it was useful in some contexts to think of a function g being indistinguishable from a constant function (in this context, the constant function corresponds to the simulator h on a level set) being equivalent to g being indistinguishable from a Bernoulli random variable.

More broadly, the notion of *computational indistinguishability* has become one of the most fundamental notions in theoretical computer science [GM82; Yao82]. We can think of the notion of computational indistinguishability as being the computational analogue of *statistical distance*. Formally:

Definition 5.1 (Computational indistinguishability)**.** Given a class $\mathcal{F}$ of distinguishers $f\colon \mathcal{X} \to \{0,1\}$, $\epsilon > 0$, and two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ on $\mathcal{X}$, we say that $\mathcal{D}_1$ and $\mathcal{D}_2$ are $(\mathcal{F}, \epsilon)$-*indistinguishable* if for all $f \in \mathcal{F}$,

$$\left| \Pr_{x \sim \mathcal{D}_1}[f(x) = 1] - \Pr_{x \sim \mathcal{D}_2}[f(x) = 1] \right| \leq \epsilon.$$

There is an important difference in settings depending on the choice of $\mathcal{F}$: namely, whether

the distinguishers are *non-uniform* or *uniform*. In the case of *non-uniform* models of computation, the distinguishers correspond to boolean circuits, whereas in the case of *uniform* models, the distinguishers corespond to polynomial time algorithms. In this section, we will only state our theorems with the class $\mathcal{F}$ corresponding to boolean circuits, and hence our results correspond to the non-uniform setting. However, we remark that the results can be extended the the uniform setting; in particular, the work of Vadhan and Zheng on which this chapter is based on [Zhe14; VZ13], state the theorems and definitions that we work with on both the non-uniform and uniform settings.

In the case of the original paper of Goldwasser and Micali, the distinguishers correspond to probabilistic polynomial-time algorithms [GM82]. Hence, the key difference is that computational indistinguishability only considers tests that are *efficient*, whereas statistical distance allows *any* test. The fact that simple computational assumptions make these two notions completely different allowed the development of secure encryption in cryptography, among many other advancements [BSW03]. This is another reason for why introducing a family $\mathcal{F}$ of distinguishers is crucial in all of the applications that we are seeing.

Given this context, the next natural information-theoretic notion that one would consider is that of entropy. Computational analogues of entropy were subsequently introduced by Yao [Yao82] and Håstad, Impagliazzo, Levin, and Luby [HILL99], the latter being known as *pseudoentropy*. The notion of pseudoentropy allowed Håstad et al. to prove the fundamental result that establishes the equivalence between pseudorandom generators and one-way functions [VZ12], which is one of the fundamental results in cryptography and complexity theory. Later, Vadhan and Zheng showed that the notion of pseudoentropy is equivalent to hardness of sampling [VZ12]. In his PhD thesis, Zheng proved a similar theorem but for average-case variants of the Håstad et al. instead, known as *pseudo-average min-entropy*, which we will refer to as *PAME* [Zhe14]. As we will now develop, the boolean case of PAME (i.e., the special case that involves only a binary alphabet) is equivalent to the dense hardcore distributions of Impagliazzo that we considered in Chapter 4. Through this relationship, we will be able to propose our generalized PAME++ theorem by employing our IHCL++ theorem from Chapter 4.

5.1 Definitions

We begin by reviewing the necessary definitions. Throughout this chapter, capitalized letters denote random variables. In particular, $\mathcal{X}$ continues to correspond to the domain, while X now denotes a random variable. In this chapter, $\mathcal{X}$ will correspond to $\{0,1\}^n$; i.e., the set of all n-bit strings. Hence, $|\mathcal{X}| = 2^n$.

The original Shannon entropy is defined as follows:

Definition 5.2 (Shannon entropy)**.** The *Shannon entropy* of a discrete random variable X is defined as

$$H(X) := \mathop{\mathbb{E}}_{x\sim X}\Big[\log\Big(\frac{1}{\Pr[X=x]}\Big)\Big].$$

Throughout this chapter, we will need to consider the conditional version of various definitions. The following is the conditional version of Shannon entropy:

Definition 5.3 (Conditional Shannon entropy)**.** The *conditional Shannon entropy* of a random variable Y given random variable Z is defined as

$$H(Y|Z) := \mathop{\mathbb{E}}_{z\sim Z}[H(Y|_{Z=z})].$$

As introduced in the beggining of this chapter, Håstad et al. introduced the following computational analogue of Shannon entropy, which they called *pseudoentropy*:

Definition 5.4 (Pseudoentropy [HILL99], informal)**.** A random variable X has *pseudoentropy at least* k if there exists a random variable Y such that

1. $H(Y) \geq k$, where $H(\cdot)$ denotes Shannon entropy;
2. X is computationally indistinguishable from Y.

One of the reasons why pseudoentropy is an interesting concept, as explained by [HILL99; VZ13], is that a random variable can have much higher pseudoentropy than its Shannon entropy. An example where this is true, for example, is the case of pseudorandom generators.

The reason why Definition 5.4 is labeled as informal is because it does not specify the class of distinguishers against which X and Y ought to be computational indistinguishable. In particular, it does not distinguish between the non-uniform and uniform settings. We can make the definition of pseudoentropy formal either by working with non-uniform distinguishers or uniform distinguishers. As we outlined in the introduction, in this chapter we will only work in the non-uniform setting. Hence, the formal definition of pseudoentropy that we work with is the following one:

Definition 5.5 (Pseudoentropy, non-uniform setting)**.** Let $\mathcal{F}$ be the class of circuits of size at most s, and let $\epsilon > 0$. We say that a random variable X has $(\mathcal{F}, \epsilon)$-*non-uniform pseudoentropy at least* k if there exists a random variable Y such that

1. $H(Y) \geq k$, where $H(\cdot)$ denotes Shannon entropy;
2. X and Y are $(\mathcal{F}, \epsilon)$-indistinguishable.

The uniform version of the notion of pseudoentropy can be found in [VZ12, Def. 210], which requires some some subtle technialities, such as the use of security parameters and sampling oracles.

Remark 5.6. While we include the term "non-uniform" in the formal definitions of pseudoentropy (Definition 5.5) and of pseudo-average-min-entropy (PAME, Definition 5.11), throughout the chapter we will be dropping the term "non-uniform" from the name, given that we do not consider the uniform setting.

A widely-used variant of Shannon's entropy is what is known as the *min-entropy*, which turns out to be the right notion to use in a lot of cryptography applications:

Definition 5.7 (Min-entropy)**.** The *min-entropy* of a random variable X is defined as

$$H_\infty(X) := \min_x \left\{ \log\left(\frac{1}{\Pr[X = x]}\right)\right\}.$$

The difference between min-entropy and Shannon entropy (Definition 5.2) is that min-entropy takes the minimum of the quantity $1/\Pr[X = x]$, whereas Shannon averages this quantity over X.

We can now obtain the analogue of Definition 5.5 (i.e., the computational analogue of Shannon entropy) but for min-entropy instead, which is known as *pseudo-min-entropy*:

Definition 5.8 (Pseudo-min-entropy [HILL99], informal)**.** A distribution X has *pseudo-min-entropy at least* k if there exists a distribution Y such that

1. $H_\infty(Y) \geq k$, where $H_\infty(\cdot)$ denotes min-entropy;
2. X is indistinguishable from Y.

Hsiao, Lu, and Reyzin considered the conditional version of pseudo-min-entropy known as *pseudo-average-min-entropy*, which we will refer to as *PAME* throughout this chapter [HLR07]. For that, we need to first define *average min-entropy*:

Definition 5.9 (Average min-entropy [DRS04])**.** For every joint distribution (X, B), the *average min-entropy of* B *given* X is defined as

$$\tilde{H}_\infty(C|X) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}[1/2^{H_\infty(C|_{X=x})}]}\right) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}[\max_a \Pr[C=a|X=x]]}\right).$$

Definition 5.9 corresponds to the conditional version of min-entropy. While there are other ways of defining the conditional version of entropies, Proposition 5.12 illustrates a very useful property of average min-entropy.

Definition 5.10 (Pseudo-average-min-entropy (PAME), informal [HLR07])**.** Let (X, B) be a joint distribution. We say that B has *PAME* $\geq k$ *given* X if there exists a distribution C jointly distributed with X such that

1. (X, B) is indistinguishable from (X, C);
2. $\tilde{H}_\infty(C|X) \geq k$, where $\tilde{H}_\infty(C|X)$ denotes the *average min-entropy* of C given X.

As we did in the case of pseudoentropy, we now formalize the notion of PAME in the non-uniform setting. Then, Definition 5.10 becomes:

Definition 5.11 (Pseudo-average min-entropy (PAME), non-uniform setting [Zhe14])**.** Let (X, B) be a joint distribution, let $\mathcal{F}$ correspond to circuits of size at most s, and let $\epsilon > 0$. We say that B has *non-uniform* $(\mathcal{F}, \epsilon)$*-pseudo-average min-entropy at least* k *given* X if there exists a random variable C jointly distributed with X such that the following holds:

1. (X, B) and (X, C) are $(\mathcal{F}, \epsilon)$-indistinguishable.
2. $\tilde{H}_\infty(C|X) \geq k$.

5.2 The PAME theorem

A known fact about (conditional) min-entropy is the following:

Proposition 5.12 ([DRS04, Proposition 4.10])**.** *For every joint distribution* (X, B)*,*

$$\tilde{H}_\infty(B|X) \geq k \iff \Pr[f(X) = B] \leq 2^{-k} \quad \forall f\colon \{0,1\}^n \to \{0,1\}.$$

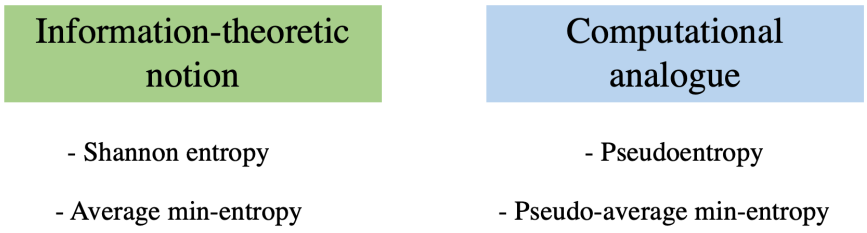

Figure 5.1: Summary of the correspondence between the information-theoretic notions and their computational analogues.

Proof. The quantity $\Pr[f(X) = B]$ is maximized when f outputs the most likely value of $B|_{X=x}$. This yields

$$\Pr[f(X) = B] = \mathop{\mathbb{E}}_{x \sim X}[\max_a \Pr[B = a | X = x]] = 2^{-H_\infty(B|X)},$$

as claimed. □

In other words, Proposition 5.12 characterizes the notion of (average) min-entropy in terms of unpredictability: If $B|X$ has high average min-entropy, then B is hard to predict from X. Moreover, Proposition 5.12 establishes an equivalence between the two notions, given that it is an if and only if statement. That is, we can characterize the notion of (average) min-entropy through the notion of unpredictability and viceversa.

One of the main theorems shown in Vadhan and Zheng [VZ12], which is the central theorem of this chapter, is the computational analogue of Proposition 5.12. Recall that pseudo-average min-entropy (PAME) is the computational analogue of average min-entropy, and hence the following theorem characterizes the notion of pseudo-average min-entropy, instead of average min-entropy, which is the notion used in Proposition 5.12. (That is, the following theorem is *exactly* the computational analogue of Proposition 5.12.)

Before we state Vadhan & Zheng's PAME theorem, we recall the definition of $(\mathcal{F}, \delta)$-weakly hard from Chapter 4, adapted to the notation of this chapter. Moreover, in what follows, we will always be working over the domain $\{0,1\}^n \times \{0,1\}$, for $n \in \mathbb{Z}$.

Remark 5.13. In the work of Vadhan & Zheng, they work over the domain $\{0,1\}^n \times \ell$, where ℓ is of order $O(\log(n))$, and hence all of their results contain a parameter ℓ [Zhe14; VZ12; VZ13]. However, for the reasons discussed in Section 5.3, we will be working in the restricted setting where $\ell = 1$. We leave it for future work to generalize our PAME++ theorem (Theorem 5.26 in Section 5.4) to larger values of ℓ, and we include a discussion at the end about our first steps in this direction.

Definition 5.14 (Hardness of prediction, non-uniform setting [Zhe14, Definition 4.13])**.** Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$, $\mathcal{F}$ any class of functions $f : \{0,1\}^n \to \{0,1\}$, and let $\delta > 0$. We say that B is *non-uniformly* $(\mathcal{F}, \delta)$*-hard to predict given* X if

$$\Pr[f(X) = B] \leq 1 - \delta.$$

If $\Pr[f(X) = B] \leq \frac{1}{2} + \epsilon$ for some $\epsilon > 0$, then we say that g is $(\mathcal{F}, \epsilon)$*-strongly hard to predict.*

We remark that this definition of hardness of prediction generalizes the notion of δ-weakly hardness that we defined in Chapter 4 (Definition 4.7). In particular, Definition 5.14 corresponds

to the notion of a function $g\colon \{0,1\}^n \to \{0,1\}$ being δ-weakly hard with respect to a distribution $\mathcal{D}$ on the domain $\mathcal{X} = \{0,1\}^n$ by setting $X = \mathcal{D}$ and $B = g(X)$.

We can now state the PAME theorem:

Theorem 5.15 (PAME theorem, informal [Zhe14])**.** *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$ and let $r > 0$. Then B has PAME $\geq r$ given X if and only if B is $(1 - 2^{-r})$-hard to predict given X.*

During this chapter, we will refer to the above theorem as the "PAME theorem". In other words, the PAME theorem states an equivalence between having high pseudo-average-min-entropy and being hard to predict.

The reason that the above theorem is informal is because we are not specifying the class of distinguishers $\mathcal{F}$. The following theorem is the formalized non-uniform version of Theorem 5.15. In that case, we separate the if and only if statement into the two directions, given that the parameters are different in the two directions.

Theorem 5.16 (PAME theorem, non-uniform setting [Zhe14, Theorem 4.15])**.** *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$, $\mathcal{F}$ any class of functions $f\colon \{0,1\}^n \to \{0,1\}$, and let $r, \epsilon > 0$. Then, the following two statements hold:*

1. *If B is non-uniformly $(\mathcal{F}_{\mathrm{poly}(n,1/\epsilon)}, 1 - 2^{-r})$-hard to predict given X, then B has non-uniform $(\mathcal{F}, \epsilon)$-PAME at least r given X.*
2. *If B has $(\mathcal{F}, \epsilon)$-PAME at least r given X, then B is non-uniformly $(\mathcal{F}, 1 - 2^{-r} - \epsilon)$-hard to predict.*

We remark that in [VZ12; Zhe14] a similar theorem is proven but using Shannon entropy instead of min-entropy and finding an equivalence to hardness of sampling (which is quantified using the Kullback-Leibler divergence) instead of to unpredictability. However, for the purposes of this chapter, we will only focus on the PAME notion, and we do not further discuss results related to pseudoentropy.

The key relationship that we are interested in exploring is the one between Impagliazzo's Hardcore Lemma (Chapter 4) and the PAME theorem. In his PhD thesis, Zheng explains how we can understand the PAME theorem as a generalization of Impagliazzo's Hardcore Lemma, and how IHCL implies a restricted version of the PAME theorem. In Section 5.3 we describe this implication, and in Section 5.4 we then use this construction coupled with our IHCL++ from Chapter 4 to obtain our PAME++ theorem.

5.3 Relationship between IHCL and the PAME theorem

The discussion in this subsection is based on Chapter 4 from Zheng's PhD thesis [Zhe14].

We begin by re-stating Impagliazzo's Hardcore Lemma (Theorem 4.9 in Chapter 4) with different notation, in order to match the information-theoretic notation of this chapter:

Theorem 5.17 (IHCL, version in [Zhe14])**.** *Let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$, $\mathcal{F}$ any class of functions $f\colon \{0,1\}^n \to \{0,1\}$, and let $\epsilon, \delta > 0$. Suppose that B is $(\mathcal{F}_{O(\log(1/\delta)/\epsilon^2)}, \delta)$-hard given X; that is, $\Pr[f(X) = B] \leq 1 - \delta$ for all $f \in \mathcal{F}$. Then there exists a joint distribution*

(X', B') that is 2δ-dense in (X, B) such that

$$\Pr[f(X') = B'] \leq \frac{1+\epsilon}{2}$$

for all $f \in \mathcal{F}_t$, where $t = s/O(\log(1/\delta)/\epsilon^2)$.

We remark that the above version of the theorem uses the 2δ density optimal parameter (first shown by Holstein) [Hol05]. Another remark is that Theorem 5.17 is stated in more generality than how we stated IHCL in Chapter 4: namely, in Chapter 4 we always assumed that the weakly hardness assumption is with respect to the uniform distribution on the domain, whereas in Theorem 5.17 above we are sampling according to the distribution X in the hardness assumption. In particular, the IHCL as stated in Theorem 5.17 corresponds to the IHCL formulation that we used in Chapter 4 by setting $X = \mathcal{U}_{\mathcal{X}}$ and $B = g(X)$ for our usual function g. (Recall that in this chapter, we are using $\mathcal{X} = \{0,1\}^n$, and that $\mathcal{U}_{\mathcal{X}}$ denotes the uniform distribution over $\mathcal{X}$. Because $\mathcal{X} = \{0,1\}^n$, we write $\mathcal{U}_{\mathcal{X}}$ as $\mathcal{U}_n$, since in this case it corresponds to the uniform distribution over n-bit strings.)

Having established this relationship, the following interpretation of IHCL provides some intuition for why IHCL is related to the PAME theorem:

> IHCL (Theorem 5.17) can be interpreted as saying the following: B cannot be predicted from X with probability $> 1 - \delta$ if and only if B is indistinguishable from a random bit on a 2δ fraction of the probability space (X, B).

We now unpack this interpretation of IHCL, explaining how it corresponds to our understanding and terminology of IHCL from Chapter 4. As we just described, we think of B as our usual function g; in particular, we can think of it as $B = g(X)$. The distribution X corresponds to some distribution $\mathcal{D}$ over the domain $\mathcal{X}$, where in the IHCL case $\mathcal{D}$ corresponds to $\mathcal{U}_{\mathcal{X}}$, and in this chapter $\mathcal{X}$ corresponds to $\{0,1\}^n$. The class $\mathcal{F}$ continues to be the set of distinguishers f. Saying that "B cannot be predicted from X with probability $> 1 - \delta$" corresponds to saying that $\Pr[f(X) = B] \leq 1 - \delta$ for all $f \in \mathcal{F}$. Lastly, the subset of the of the domain in which B is indistinguishable from a random bit corresponds to what we called the *hardcore set* in Chapter 4. Being indistinguishable from a random bit corresponds to the notion of strong hardness from Chapter 4. This matches exactly the intuitive explanation that we provided for the conclusion of the IHCL in Chapter 4: we described how inside the hardcore set, function g behaves like a random boolean function, because the distinguishers cannot guess the outputs of g with probability larger than $1/2$ (plus ϵ slack). This is the same idea as saying that g is behaving like a random bit.

Before we describe how Impagliazzo's Hardcore Lemma implies a restricted version of the PAME theorem, a preliminary question that arises is the following: How can we speak of the equivalence of these two theorems if the PAME theorem is an "if and only if" statement while IHCL is not? The reason behind this discrepancy is that IHCL could be stated as an if and only if statement; however, one of the directions is never included in the IHCL statement because it follows trivially from definition. This corresponds to the implication: If there exists a hardcore set S of density 2δ over which g is strongly hard, then g is δ-weakly hard on average.

Where is the notion of pseudo-average min-entropy (PAME) in the IHCL statement? The next natural question is: where does the notion of pseudo-average-min-entropy appear in the IHCL statement? By Proposition 5.12, we know that when $\mathcal{F}$ corresponds to the set of all functions $f : \{0,1\}^n \to \{0,1\}$, then $\Pr[f(X) = b] \leq 2^{-k}$ is equivalent to $\tilde{H}_\infty(B|X) \geq k$. In this case,

$$k = \log\Big(\frac{1}{1-\delta}\Big),$$

since then

$$2^{-\log\Big(\frac{1}{1-\delta}\Big)} = 2^{\log(1-\delta)} = 1-\delta.$$

Therefore, the assumption that B is $(\mathcal{F}, \delta)$-hard to predict given X implies that

$$\tilde{H}_\infty(B|X) \geq \log\Big(\frac{1}{1-\delta}\Big).$$

After having gained some intuition on the underlying relationship between IHCL and the PAME theorem we will now prove the following, which corresponds to one of the two directions stated in Theorem 5.16 (and the direction we are interested in for our purposes):

Theorem 5.18 (PAME theorem, restricted version, informal)**.** *Let $(\mathcal{U}_n, g(\mathcal{U}_n))$ be a joint distribution on $\{0,1\}^n \times \{0,1\}$. Then B has PAME $\geq r$ given X if and only if B is $(1-2^{-r})$-hard to predict given X.*

Remark 5.19. We will now show how the ICHL (Theorem 4.9 from Chapter 4) implies PAME (Theorem 5.16) when setting $X = \mathcal{U}_n$, $B = g(X)$. As we just described, the IHCL stated in Theorem 5.17 is more general, and so the implication that IHCL implies PAME still holds for general (X, B). However, we restrict it to the case $X = \mathcal{U}_n$, $B = g(X)$ because our goal is to use the IHCL++ theorem from Chapter 4 to obtain PAME++. Still, the proof that we present is not restricted to the case $X = \mathcal{U}_n$, $B = g(X)$, and so in future work we could also begin by writing IHCL++ for (B, X) instead.

As we have discussed, one direction of Theorem 5.18 is trivial. Therefore, following the ideas outlined in [Zhe14, Ch. 4], we prove the following, which corresponds to the second statement in Theorem 5.16 above:

Theorem 5.20 (Restricted PAME, one direction, non-uniform setting)**.** *Let $\mathcal{F}$ be any class of functions $f\colon \{0,1\}^n \to \{0,1\}$, let $\epsilon, \delta > 0$, and let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$ where $X = \mathcal{U}_n$ and $B = g(X)$ for some $(\mathcal{F}_{\mathrm{poly}(n,1/\epsilon)}, \delta)$-weakly hard function g. Then, B has non-uniform $(\mathcal{F}, \epsilon)$-PAME $\geq \log(1/(1-\delta))$ given X.*

Proof. The intuition for the proof is as follows: Because function g is δ-weakly hard by assumption, we can apply IHCL to g to obtain a hardcore set H in $\mathcal{X} = \{0,1\}^n$. Then, we define a distribution (X, C) based on the hardcore set H, which we can show has enough average min-entropy. Intuitively, this is a sensible approach precisely because a hardcore set is related to the notion of *unpredictability*. That is, within the hardcore set, g behaves like a random function. It is then natural to expect this unpredictability to yield high min-entropy when we define a distribution based on the hardcore set.

By assumption, g is $(\mathcal{F}_{\mathrm{poly}(n,1/\epsilon)}, \delta)$-weakly hard. By definition, this implies that

$$\Pr_{x\sim X}[f(x) = g(x)] \leq 1 - \delta \quad \forall f \in \mathcal{F}_{\mathrm{poly}(n,1/\epsilon)},$$

given that $B = g(X)$. Then, by IHCL (Theorem 4.9), this implies that there exists a hardcore set $H \subseteq \{0,1\}^n$ with $|H| \geq 2\delta \cdot 2^n$ (i.e., with density $\geq 2\delta$, given that here $|\mathcal{X}| = 2^n$) with respect to the class $\mathcal{F}_{\mathrm{poly}(n,1/\epsilon)}$.

Construction of a distribution with high PAME given a hardcore set. Next, we define (X, C) using the hardcore set H as follows. Given $X = x$,

$$C(x) := \begin{cases} \{0,1\} \text{ each with probability } 1/2, & \text{if } x \in H, \\ g(x), & \text{if } x \notin H. \end{cases} \tag{5.21}$$

In other words, the first case corresponds to returning a uniform random bit. Then, we claim the following:

Claim 5.22. $\tilde{H}_\infty(C|X) \geq \log\left(\frac{1}{1-\delta}\right)$.

Proof. In order to prove this, we use the original definition of average min-entropy (Definition 5.9). Namely, recall that

$$\tilde{H}_\infty(C|X) := \log\left(\frac{1}{\mathbb{E}_{x\sim X}[1/2^{H_\infty(C|_{X=x})}]}\right).$$

When $x \in H$, then $H_\infty(C|_{X=x}) = 1$, given that $C(x)$ is a random bit (by 5.21). When $x \notin H$, then $H_\infty(C|_{X=x}) = 0$, given that $C(x) = g(x)$ (by 5.21), and g is a deterministic function by assumption.

Then,

$$\mathop{\mathbb{E}}_{x\sim X}\left[1/2^{H_\infty(C|X=x)}\right] = 2\delta \cdot \frac{1}{2} + (1 - 2\delta) = 1 - \delta.$$

Therefore,

$$\tilde{H}_\infty(C|X) = \log\left(\frac{1}{\mathbb{E}_{x\sim X}[1/2^{H_\infty(C|_{X=x})}]}\right) = \log\left(\frac{1}{1-\delta}\right),$$

as required.

□

Next, we show that:

Claim 5.23. *(X, B) and (X, C) are $(\mathcal{F}, \epsilon)$-indistinguishable.*

Proof. As in Claim 5.22, we consider the cases $x \in H$ and $x \notin H$ separately. When $x \in H$, then by definition of C, C is a uniform random bit. Since B is δ-weakly hard by assumption, the IHCL implies that B is also $(\mathcal{F}_t, \epsilon)$-strongly hard over H. Therefore, by definition of strong hardness, (X, B) and (X, C) are $(\mathcal{F}_t, \epsilon)$-indistinguishable inside H.

When $x \notin H$, by definition of C it follows that $C(x) = g(x)$. Since $B = g(X)$ by assumption, C equals B outside of H, and hence trivially (X, B) and (X, C) are $(\mathcal{F}_t, \epsilon)$-indistinguishable since they are, in fact, equal to each other.

Hence, since (X, B) and (X, C) are $(\mathcal{F}_t, \epsilon)$-indistinguishable both inside and outside of the hardcore set H, it follows that (X, B) and (X, C) are $(\mathcal{F}_t, \epsilon)$-indistinguishable over the entire domain, as we wanted to show. □

By definition of PAME (Definition 5.10), Claims 5.22 and 5.23 together imply the following:

Corollary 5.24. *B has (non-uniform) $(\mathcal{F}_t, \epsilon)$-PAME $\geq \log(1/(1-\delta))$ given X.*

Proof. By definition of $(\mathcal{F}_t, \epsilon)$-PAME, we need to show that there exists a random variable C jointly distributed with X such that (1) $\tilde{H}(C|X) \geq k$, and (2) (X, B) and (X, C) are $(\mathcal{F}_t, \epsilon)$-indistinguishable. By letting C be the random variable defined in Equation 5.21, it follows that Condition (1) is fulfilled by Claim 5.22 and Condition (2) is fulfilled by Claim 5.23. □

This concludes the proof of Theorem 5.20. □

5.4 Our proposed PAME++

Given the relationship between IHCL and PAME that we just explored in Section 5.3, we return to the main underlying question of this thesis: Given that a multiaccurate predictor implies IHCL (as shown in Chapter 4), and given that we just showed that IHCL implies (a restricted version of) the PAME theorem, if we begin with a multicalibrated predictor instead, what stronger (++) version of PAME do we obtain? To do so, we will use our IHCL++ theorem from Chapter 4, and plug it into the reduction from IHCL to PAME that we just showed.

Given our goal, let us recall our proposed IHCL++ (set version) from Chapter 4 (Theorem 4.11):

Theorem 5.25 (IHCL++, Set version)**.** *Let $\mathcal{X}$ be a finite domain, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, let $g\colon \mathcal{X} \to \{0,1\}$ be an arbitrary function, let $\epsilon, \gamma > 0$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ which satisfies that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$ and such that $|\mathcal{F}| \leq \frac{1}{10}\exp(2|\mathcal{X}|\epsilon^2 k_p^2)$, there exists a set $H_p \subseteq P$ of density $|H_p|/|P| \geq 2k_p$ such that H_p is an $(\mathcal{F}, \epsilon/k_p)$-hardcore set for g. That is,*

$$\forall f \in \mathcal{F}, \quad \Pr_{x \in H_p}[f(x) = g(x)] \leq \frac{1}{2} + \frac{\epsilon}{k_p}.$$

Recall from Chapter 4 that η_p corresponds to the size parameter of $P \in \mathcal{P}$ and k_p corresponds to the balance of g on P. The idea on how to go from IHCL++ to PAME++ is that, for every hard-core set H_p, we can build a joint distribution with high PAME in the larger set where the hardcore set is contained. This is what enabled us to prove Theorem 5.20 as well, now replicated to each set $P \in \mathcal{P}$. Following this intuition, we propose the following original PAME++ theorem:

Theorem 5.26 (PAME++)**.** *Let $\mathcal{F}$ be any class of functions, let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$ where $X = \mathcal{U}_n$ and $B = g(X)$ for an arbitrary boolean function g, and let $\epsilon, \gamma > 0$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ which satisfies that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, $B|_P$ has non-uniform $(\mathcal{F}, \epsilon/k_p)$-PAME at least $\log(1/(1-k_p)) = H_\infty(g(\mathcal{U}_p))$ given $X|_P$, where $B|_P$ denotes the restriction of B on P and $X|_P$ the restriction of X on P.*

Remark 5.27. For simplicity in the argument, we use the set version of IHCL++. However, PAME++ can be shown directly using the measure version of IHCL++, which is why we drop the assumption parameters on $|\mathcal{F}|$ in the statement of PAME++.

Proof. We follow the proof strategy for Theorem 5.20. Recall that $v_p = \mathbb{E}_{x \in P}[g(x)]$, and $k_p = \min\{v_p, 1 - v_p\}$. By the ICHL++ (Theorem 5.25), we know that there exists a partition $\mathcal{P}$ of $\mathcal{X}$ satisfying that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, there exists a set $H_p \subseteq P$ of density $|H_p|/|P| \geq 2k_p$ such that, for all $f \in \mathcal{F}$,

$$\Pr_{x \sim H_p}[f(x) = g(x)] \leq 1/2 + \frac{\epsilon}{k_p}.$$

For each $P \in \mathcal{P}$, we define a joint distribution (X_p, C_p) on the domain $P \times \{0,1\}$ as follows. Given $X_p = x$,

$$C_p(x) = \begin{cases} \{0,1\} \text{ each with } \Pr = 1/2, & \text{if } x \in H_p, \\ g(x), & \text{if } x \notin H_p. \end{cases}$$

Then, as in the proof of Theorem 5.20, we show the following analogous two claims:

Claim 5.28. $\tilde{H}_\infty(C_p|X_p) \geq \log\left(\frac{1}{1-k_p}\right)$.

Proof. This follows analogously to the proof of Claim 5.22, by using the restriction C_p instead of C and X_p instead of X. □

Claim 5.29. *(X_p, C_p) and $(X|_P, B|_P)$ are $(\mathcal{F}, \epsilon/k_p)$-indistinguishable.*

Proof. This follows directly by the proof of Claim 5.23, by using C_p instead of C, $X_p = X|_P$ instead of X, and $B|_P$ instead of B. That is, if $x \in H_p$, then by IHCL++ we know that g is $(\mathcal{F}, \epsilon/k_p)$-strongly hard on H_p. Therefore, (X_p, C_p) and $(X|_P, B|_P)$ are $(\mathcal{F}, \epsilon/k_p)$-indistinguishable. When $x \notin H_p$, the two joint distributions are equal to each other, and hence they are also $(\mathcal{F}, \epsilon/k_p)$-indistinguishable. □

We remark that X_p and $X|_P$ are equivalent by definition. Together, by the definition of PAME (Definition 5.11, Claims 5.28 and 5.29 imply the following:

Corollary 5.30. *$B|_P$ has (non-uniform) $\left(\mathcal{F}_t, \epsilon/k_p\right)$-PAME $\geq \log(1/(1-k_p))$ given $X|_P$.*

Having shown Claim 5.30, we now justify the simplifications to the expression $\log(1/(1-k_p))$. First, we show that

$$\log\left(\frac{1}{1-k_p}\right) = \log\left(\frac{1}{2^{-H_\infty(g(U_p))}}\right).$$

We recall that $X = \mathcal{U}_n$, where $\mathcal{U}_n$ denotes the uniform distribution over $\mathcal{X} = \{0,1\}^n$ (i.e., over n-bit strings). Hence, within a level set P, we have that $X|_P = \mathcal{U}_P$, where $\mathcal{U}_P$ denotes the uniform distribution over P. Then,

$$\begin{aligned} 1 - k_p &= 1 - \min\{\mathbb{E}_{x\in P}[g(x)], 1 - \mathbb{E}_{x\in P}[g(x)]\} = 1 - \min\{\Pr[g(U_p) = 1], \Pr[g(U_p) = 0]\} \\ &= \max\{\Pr[g(U_p) = 0], \Pr[g(U_p) = 1]\} = 2^{-H_\infty(g(U_p))}, \end{aligned}$$

where the last equality follows from the definition of min-entropy (Definition 5.7). Then,

$$\log\left(\frac{1}{2^{-H_\infty(g(U_p))}}\right) = \log\left(2^{H_\infty(g(U_p))}\right) = H_\infty(g(U_p)).$$

This matches the expression stated in Theorem 5.26. □

We remark that although we are proving PAME++ using IHCL++, recall that we used an approximate multicalibrated partition $\mathcal{P}$ to prove IHCL++. Therefore, the partition $\mathcal{P}$ that proves the PAME++ theorem is an approximate multicalibrated partition, which is the fundamental idea behind the theorem.

5.4.1 Recovering the original PAME++ from PAME

Similar to how we proceeded in Chapter 4, we now show how to recover the original PAME theorem using PAME++. As in the case of IHCL++, the key idea is to "glue" together the sets $P \in \mathcal{P}$ that have enough size and are balanced enough. Namely, such that η_p and k_p are larger than some threshold. Recall the definition of a "good" $P \in \mathcal{P}$ from Chapter 4: we say that a set $P \in \mathcal{P}$ is (γ,τ)-good for some $\gamma, \tau > 0$ if $\eta_p \geq \gamma$ and $k_p \geq \tau$. Recall that in the original PAME theorem (Theorem 5.20) we are assuming that g is δ-weakly hard, unlike in the PAME++ statement.

Proof of PAME using PAME++. Let $\mathcal{F}, \mathcal{X}, \epsilon, \delta$ be the assumption parameters in PAME. We define the parameters $\epsilon' := \epsilon^2\delta$, $\gamma := \epsilon\epsilon'$, and invoke the PAME++ theorem with these parameters ϵ', γ. By PAME++, we obtain a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon'^4\gamma)\cdot\log(|\mathcal{X}|/\epsilon'))$, $q = O(1/\epsilon'^2)$, $k = O(1/\epsilon')$ such that, for each $P \in \mathcal{P}$ where $\eta_p \geq \gamma = \epsilon\epsilon'$, $B|_P$ has non-uniform $(\mathcal{F}, \epsilon'/k_p)$-PAME at least $\log(1/(1-k_p)) = H_\infty(g(\mathcal{U}_P))$ given $X|_P$.

Let $\tau := \epsilon\delta$. By Proposition 4.12 from Chapter 4 we know that

$$\mathop{\mathbb{E}}_{P\sim\mathcal{P}}[k_p \cdot \mathbb{1}_G(P)] \geq \delta \cdot (1 - O(\epsilon)).$$

In order to prove the PAME theorem, we need to show that B has $(\mathcal{F}, \alpha)$-PAME at least equal to $\log(1/(1-\delta))$ given X. To do so, by definition of PAME, we need to show that there exists a distribution C jointly distributed with X such that (1) $\tilde{H}_\infty(C|X) \geq \log(1/(1-\delta))$, and (2) (X, B) and (X, C) are $(\mathcal{F}, \alpha)$-indistinguishable.

As we did in Chapter 4, we construct such a C by doing "gluing together" the distributions C_P obtained from invoking the PAME++ theorem such that P is *good*. Namely, for each $P \in \mathcal{P}$, let C_P be distributed as $\mathcal{C}_P$. Then, we define the distribution $\mathcal{C}$ on $\mathcal{X}$ as

$$\mathcal{C}(x) = \mathcal{C}_P(x) \cdot \mathbb{1}_G(P).$$

Let C correspond to the random variable distributed as $\mathcal{C}$; i.e., $C \sim \mathcal{C}$. By the PAME++ theorem, we know that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$,

$$\tilde{H}_\infty(C_p|X_p) \geq \log\Big(\frac{1}{1-k_p}\Big).$$

Hence, by the definition of C and by applying the bound $\mathbb{E}_{P\sim\mathcal{P}}[k_p \cdot \mathbb{1}_G(P)] \geq \delta \cdot (1 - O(\epsilon))$, it follows that

$$\tilde{H}_\infty(C|X) \geq \log\Big(\frac{1}{1-\delta\cdot(1-O(\epsilon))}\Big).$$

Similarly, by the PAME++ theorem we know that every joint distribution (X_p, C_p) is $(\mathcal{F}, \epsilon'/k_p)$-indistinguishable from $(X|_P, B|_P)$ for each $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$. Therefore, since in Chapter 4 we showed that

$$\mathop{\mathbb{E}}_{P\sim\mathcal{P}}\left[\frac{\epsilon'}{k_p}\right] \leq \epsilon'/\tau = \epsilon,$$

it follows that, by definition of C, distributions (X, C) and (X, B) are $(\mathcal{F}, \epsilon)$-indistinguishable.

Putting the two facts together, by definition of PAME (Definition 5.11), it follows that B has non-uniform $(\mathcal{F}, \epsilon)$-PAME at least

$$\log\Big(\frac{1}{1-\delta\cdot(1-O(\epsilon))}\Big).$$

As we explained in Chapter 4, it is possible to modify the distribution $\mathcal{C}$ to achieve the lower bound $\log(1/(1-\delta))$ on the PAME of B while changing the indistinguishability parameter between (X, C) and (X, B) by at most $O(\epsilon)$. Hence, this proves the PAME theorem (Theorem 5.16). □

6

Dense Model Theorem

A theorem of Green, Tao, and Ziegler can be stated (roughly) as follows: if R is a pseudorandom set, and D is a dense subset of R, then D may be modeled by a set M that is dense in the entire domain such that D and M are indistinguishable.

Reingold et al. [RTTV08]

The Dense Model Theorem (DMT) is a result from additive combinatorics that states the following [TTV09]: Let R be a pseudorandom subset of a set $\mathcal{X}$, which can be very sparse, and let $D \subseteq R$ such that $|D| \leq \delta|R|$. Then, there exists a model set $M \subseteq X$ such that $|M| \geq \delta|X|$ (i.e., M is *dense*) and M is indistinguishable from D. This is why M is called a *dense model* for D. When formalizing these definitions below, we will see that the model for D corresponds to a measure.

The original proof follows a potential energy argument with via iterative partitioning, similar to the proofs of the statements that we described in Chapter 3. In [RTTV08], they prove the DMT via this same approach, and also provide a new proof of Impagliazzo's Hardcore Lemma using the iterative partitioning argument as well.

One of the key motivations for the Dense Model Theorem is that it is one of the crucial proof components used in Green and Tao's famous result that there exist arbitrarily long arithmetic progressions of primes [GT08]. In their setting, $\mathcal{X}$ corresponds to $\mathbb{Z}$, R corresponds to a "pseudorandom" set of integers, and D is a subset of constant density within R. Then, by using the Dense Model Theorem and Szemerédi's theorem for arithmetic progressions, they show that D contains arbitrarily long arithmetic progressions. Then, they show that there is a set R of integers that is pseudorandom and such that the primes have constant density inside R, which allows them to conclude the proof of the Green-Tao theorem. We explore some of these connections in Chapter 6, but for now, this provides an important motivational reason for studying the Dense Model Theorem. We remark that in this chapter we use a slightly different formulation of the Dense Model Theorem, but we will describe how it implies the version of the DMT that we just presented in the opening of this chapter. A more general Dense Model Theorem was later proven by Tao and Ziegler [TZ08], which generalized it to other domains. Gowers [Gow10] and Reingold et al. [RTTV08] later provided a simplified proof using an iterative partitioning argument.

6.1 Definitions

In order to state the DMT, we will need the following definitions.

As usual, we use $\mathcal{D}$ to denote a probability distribution over a finite domain $\mathcal{X}$ and $\mathcal{F}$ to denote a class of boolean distinguisher tests f on $\mathcal{X}$. As in the rest of this thesis, $\mathcal{F}$ is always assumed to contain the constant functions and is closed under complement (Remarks 2.2, 2.3).

Definition 6.1. Given a function $f \in \mathcal{F}$ and a probability distribution $\mathcal{D}$ over $\mathcal{X}$, we define $f[\mathcal{D}]$ as

$$f[\mathcal{D}] := \Pr_{x \sim \mathcal{D}}[f(x) = 1].$$

Definition 6.2 (Indistinguishable distributions)**.** Given a class $\mathcal{F}$ of functions $f \colon \mathcal{X} \to \{0,1\}$ and two distributions $\mathcal{D}_1$, $\mathcal{D}_2$ on $\mathcal{X}$, we say that $\mathcal{D}_1$ and $\mathcal{D}_2$ are $(\mathcal{F}, \epsilon)$-*indistinguishable* if, for all $f \in \mathcal{F}$,

$$|f[\mathcal{D}_1] - f[\mathcal{D}_2]| < \epsilon.$$

Equivalently (by Definition 6.1), if

$$\left| \Pr_{x \sim \mathcal{D}_1}[f(x) = 1] - \Pr_{x \sim \mathcal{D}_2}[f(x) = 1 \right| \leq \epsilon.$$

Definition 6.3 (Pseudorandom)**.** A distribution $\mathcal{D}$ on $\mathcal{X}$ is $(\mathcal{F}, \epsilon)$-*pseudorandom* if it is $(\mathcal{F}, \epsilon)$-indistinguishable form the uniform distribution on $\mathcal{X}$ (i.e., from $\mathcal{U}_{\mathcal{X}}$).

In Chapter 4, we defined the density of a measure and of a set. However, we implicitly defined the two terms with respect to the distribution $\mathcal{D} = \mathcal{U}_{\mathcal{X}}$ over $\mathcal{X}$. We now re-state Definitions 4.5 and 4.6 for an arbitrary distribution $\mathcal{D}$ over $\mathcal{X}$.

Definition 6.4 (Density of a measure)**.** Given a distribution $\mathcal{D}$ over $\mathcal{X}$, a *measure* μ is a map from $\mathcal{X}$ to $[0,1]$ with density

$$d(\mu) = \sum_{x \in \mathcal{X}} \mu(x)\mathcal{D}(x) = \mathop{\mathbb{E}}_{x \sim \mathcal{D}}[\mu(x)].$$

A measure μ of positive density induces a distribution

$$D_\mu(x) = \frac{\mu(x)\mathcal{D}(x)}{d(\mu)}.$$

Recall from Chapter 4 that every set $S \subseteq \mathcal{X}$ has a corresponding measure given by the associated characteristic function χ_S. Then, the density of a subset S is just the probability mass endowed to S by distribution $\mathcal{D}$. Hence a set S also induces a distribution, denoted by D_S.

Definition 6.5 (Density of a set)**.** Given a distribution $\mathcal{D}$ over $\mathcal{X}$, the *density* of a set $S \subseteq$ in $\mathcal{D}$ is given by

$$d(S) = \sum_{x \in S} \mathcal{D}(x) = \Pr_{x \sim \mathcal{D}}[x \in S].$$

Definition 6.6. Given a class $\mathcal{F}$ of functions $f : \mathcal{X} \to \{0,1\}$ and a measure μ on $\mathcal{X}$,

$$f[\mu] := f[D_\mu],$$

where D_μ denotes the distribution induced by μ.

Hence, for a set $S \subseteq \mathcal{X}$,

$$f[S] = \Pr[f(x) = 1 \mid x \in S],$$

since we identify S with the conditional distribution given $x \in S$. We can similar refer to $f[\mathcal{X}]$.

Definition 6.7 (ϵ-model [Imp09])**.** Given a class $\mathcal{F}$ of functions on $\mathcal{X}$, a measure μ on $\mathcal{X}$, a set $S \subseteq \mathcal{X}$, and $\epsilon > 0$, we say that μ is an $(\mathcal{F}, \epsilon)$-*model* for S if D_S and D_μ are $(\mathcal{F}, \epsilon)$-indistinguishable.

If S has density δ in $\mathcal{D}$, then for every $f : \mathcal{X} \to \{0, 1\}$ the following holds:

$$\Pr_{x \sim \mathcal{D}}[f(x) = 1] \geq \Pr_{x \sim \mathcal{D}}[x \in S] \cdot \Pr[f(x) = 1 \mid x \in S] = \delta \cdot \Pr_{x \sim D_S}[f(x) = 1].$$

This motivates the definition of *pseudodensity*, which thus relaxes the notion of density:

Definition 6.8 (Pseudodensity [Imp09])**.** Given a class $\mathcal{F}$ of functions on $\mathcal{X}$, a distribution $\mathcal{D}$ on $\mathcal{X}$, a set $S \subseteq \mathcal{X}$, and $\epsilon, \delta > 0$, we say that S is $(\mathcal{F}, \epsilon, \delta)$-*pseudodense* in $\mathcal{D}$ if for all $f \in \mathcal{F}$,

$$\Pr_{x \sim \mathcal{D}}[f(x) = 1] \geq \delta \cdot \Pr_{x \sim D_S}[f(x) = 1] - \epsilon.$$

Since pseudodensity is a relaxation of density, we can think of pseudodensity as saying that the distinguishers $f \in \mathcal{F}$ cannot tell whether S is small (i.e., has low density). As we showed above, if S is δ-dense, then S is (ϵ, δ)-pseudodense. But another way for S to be pseudodense is the following: If there is a measure μ that is indistinguishable from S by $\mathcal{F}$ and such that μ is δ-dense [Lee17]. The Dense Model Theorem precisely states that we can always find such a μ. Hence, the interesting case is when $d(S) \ll d(\mu)$, because then the DMT states that, even if S is very small, we can find a measure of large density that is indistinguishable from S to $\mathcal{F}$.

Recall that in Chapter 4, the assumption and the conclusion of IHCL dealt with a different class of distinguishers. Namely, in IHCL, the δ-weakly hardness of g is with respect to the enlarged class $\mathcal{F}_s$ (which includes all functions with complexity at most s relative to $\mathcal{F}$), whereas the strong hardness conclusion of g is with respect to $\mathcal{F}$. In the case of the Dense Model Theorem, we have a similar phenomenon, where the pseudodensity assumption of a set S is with respect an enlarged class of distinguishers, whereas the conclusion is with respect to the original class $\mathcal{F}$. In this case, the enlarged class of distinguishers is different than that in Chapter 4, which is defined as follows:

The DMT version that we use requires the following notation:

Definition 6.9 ([Imp09])**.** Given a class $\mathcal{F}$ of functions $f \colon \mathcal{X} \to \{0, 1\}$ and an integer m, the class $\mathcal{F}^m$ is defined as

$$\mathcal{F}^m := \{\mathrm{Maj}_j(f_1, \ldots, f_j) \mid f_i \in \mathcal{F}; 1 \leq j \leq m\},$$

where $\mathrm{Maj}_j(g_1, \ldots, g_j)$ denotes the function that outputs 1 if at least $\lceil j/2 \rceil$ of the g_i's output 1.

"Maj" stands for "majority", since the definition of Maj_j corresponds to taking the majority vote of j functions. Therefore, $\mathcal{F}^m$ corresponds to the class of distinguishers that are obtained by taking the majority vote for up to m functions from the class $\mathcal{F}$.

We can now state the Dense Model Theorem.

Theorem 6.10 (DMT, [Imp09, Thm. 3], [RTTV08])**.** *Let $\mathcal{X}$ be a finite domain, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, and $\epsilon, \delta > 0$. Then there exists an $m = \text{poly}(1/\epsilon, 1/\delta)$ such that if any set $S \subseteq \mathcal{X}$ is $(\mathcal{F}^m, \epsilon, \delta)$-pseudodense in $\mathcal{U}_\mathcal{X}$, then there exists a $(\delta - O(\epsilon))$-dense measure μ such that μ is an $(\mathcal{F}, O(\epsilon/\delta))$-model for S.*

(This formulation of the DMT implies the one that we used to describe DMT informally at the beginning of this chapter by the definition of pseudorandomness; see [Imp08].)

Therefore, in order to prove the DMT, we need to construct a measure μ such that:

1. (Density) μ has density $\delta - O(\epsilon)$.
2. (Indistinguishability) D_μ and D_S are $O(\epsilon/\delta)$-indistinguishable with respect to $\mathcal{F}$ (by definition of a model; Definition 6.7).

Similarities to Impagliazzo's Hardcore Lemma. The Dense Model Theorem shares many similarities to Impagliazzo's Hardcore Lemma (Chapter 4). We highlight some of their parallels, which also apply to their ++ counter-parts, as we will see with our DMT++ statement.

1. **The assumption.** In IHCL, we assume that the function g is (F_s, δ)-weakly hard. In DMT, we assume that the S is $(\mathcal{F}^m, \epsilon, \delta)$-pseudodense.
2. **The density.** In IHCL, we want to ensure that the hardcore measure μ is dense enough; more concretely, 2δ-dense. In DMT, we want to ensure that the model μ is dense enough; more concretely, $(\delta - O(\epsilon))$-dense.
3. **Indistinguishability.** In IHCL, we need to ensure that g is strongly hard when sampling according to μ, which means that g behaves like a random function. In DMT, we want the model μ to be indistinguishable from S.

6.2 Proving the DMT using IHCL

As we would guess from the previous similarities, we would hope to prove the DMT using the IHCL. The reason we are interested in proving the DMT through the IHCL is because, per the goal of this thesis, we want to find the right generalization for DMT++. Given our IHCL++ theorem in Chapter 4, (Theorem 4.11), we could then use the ICHL-DMT correspondence to establish an equivalent IHCL++-DMT++ correspondence.

Although we actually obtain our DMT++ theorem directly from a multicalibrated partition, rather than through IHCL, we used ideas from the proof that IHCL implies DMT in our DMT++ statement. Moreover, this implication is not formalized in the literature, although it is sketched in several drafts and workshop notes by Impagliazzo. For these two reasons, we believe it is of interest to formalize how IHCL implies the DMT.

This formalization is based on the proofs provided in [Imp08; Imp09; Lee17; GIK12].

Intuitive idea. The general idea behind the reduction is as follows: in order to be able to apply IHCL in the DMT setting, we need a function g, given that there are no functions in the DMT

statement (besides the distinguishers); only sets. The natural way to obtain a function is to take g to be the characteristic function of the set S (i.e., $g(x)$ returns 1 if and only if $x \in S$, so $g := \chi_S$).

Then, we would show that because S is (ϵ, δ)-dense by assumption, this implies that g is δ-weakly hard. We can then apply IHCL to g to obtain a hardcore measure μ. Lastly, we build a dense model μ' from μ: the strong hardness condition of g on μ guarantees that μ' and S are indistinguishable, and the fact that μ is δ-dense guarantees that μ' is dense enough.

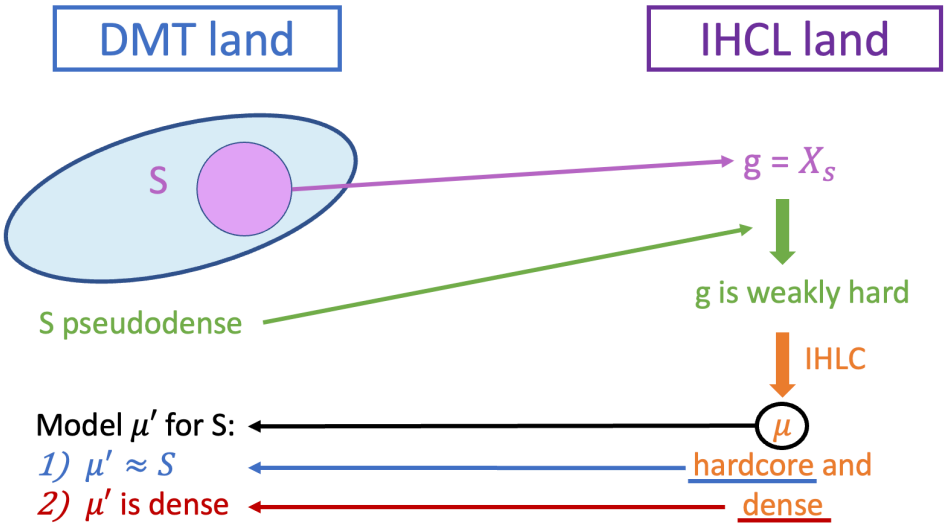

Figure 6.1: Using IHCL to prove DMT, simplified.

Magnifying S. While this is the right intuitive idea behind the reduction, a problem that arises is that we are usually interested in the setting where S is very small. That is, as we explained above, the interesting applications of DMT occur when $d(S) \ll d(\mu)$, yet μ is a model for S. But if S is very small, then the characteristic function of S can be approximated by the constant 0 function. In order to avoid this, the key idea by Impagliazzo in [Imp08; Imp09; Lee17; GIK12] is to magnify S so that we ensure that we sample from S with some constant probability δ'. In particular, we will need to set $\delta' = \delta/(1+\delta)$.

To do so, we first augment the original domain $\mathcal{X}$ as follows. Let $\mathcal{X}$ be the initial domain and S the $(\mathcal{F}^m, \epsilon, \delta)$-pseudodense set in $\mathcal{X}$ given in the DMT++ statement. From $\mathcal{X}$, we build the sets

$$V_S = \{(1, x) \mid x \in S\},$$

$$V_U = \{(0, x) \mid x \in \mathcal{X} \setminus S\}.$$

We then construct the augmented domain V as

$$V = V_S \cup V_U.$$

We remark that this is not what is magnifying S, given that the elements in V correspond exactly to the elements in U, except that we have "tagged" them with a 0 or 1 bit in order to indicate whether they are coming from S or from $\mathcal{X}$. This "tag" is just useful for defining the class of

distinguishers $\mathcal{F}$ and the function g. However, the number of elements in the domain V is exactly equal to the number of elements in the domain $\mathcal{X}$. The magnification of S instead comes from the distribution that we will specify on V.

We think of the class $\mathcal{F}$ as the set of functions $f\colon V \to \{0,1\}$ such that they ignore the first bit of the input. That is, given an arbitrary class of distinguishers $\mathcal{F}'$ on $\mathcal{X}$, we build the class $\mathcal{F}$ on V as follows. For each $f' \in \mathcal{F}'$ and $x = b \circ x' \in V$, where $x' \in \mathcal{X}$ and $b \in \{0,1\}$, we add the corresponding function f to $\mathcal{F}$: $f(x) := f'(x')$. (The symbol $\circ$ denotes the composition of two strings.)

Next, we define the following function g:

$$g(b, x) = b.$$

That is, g distinguishes which part of V does x come from: V_S or V_U? Since $b = 1$ if $x \in S$ and $b = 0$ if $x \in U$, the output of g corresponds indicates exactly the set that x belongs to.

The key idea behind the proof that IHCL implies the DMT, which we will now develop, is as follows: because S has large pseudo-density and the distinguishers ignore the first bit, they will not be able to distinguish where the input comes from (V_S or V_U), and so they "find it hard" to compute g. Formally, we first show that g is indeed weakly hard to compute on average, in the IHCL sense.

Remark 6.11. In Chapter 4, we always applied the definition of δ-weakly hardness with respect to the uniform distribution $\mathcal{X}$. In this chapter, we will use the notion of δ-weakly hardness with respect to an arbitrary distribution $\mathcal{D}$ over $\mathcal{X}$. That is, we are interested the quantity $\Pr_{x\sim\mathcal{D}}[f(x) = g(x)]$ rather than the particular case of $\Pr_{x\sim\mathcal{U}_\mathcal{X}}[f(x) = g(x)]$.

Step 1: $g(b, x) = b$ is weakly hard to compute. Let δ correspond to the parameter in the DMT statement (Theorem 6.10). We define the following distribution $\mathcal{W}$ on V generated by the following process:

- With probability $\delta' = \dfrac{\delta}{1+\delta}$, sample uniformly from V_S.
- With probability $(1 - \delta')$, sample uniformly from V_U.

That is, first we choose one of V_S or V_U according to the δ' parameter. Then, conditioned on being inside either V_S or V_U, we pick an element x uniformly. This definition of $\mathcal{W}$ is what "magnifies S" in the sense that we described. That is, the definition of $\mathcal{W}$ ensures precisely that the probability that the element x picked by $\mathcal{W}$ comes from S is δ'; hence, we are sampling from S with constant probability δ' [Imp09].

Claim 6.12. *The function g is $(\mathcal{F}^m, \delta' - (1-\delta')\epsilon)$-weakly hard with respect to distribution $\mathcal{W}$.*

Proof. For the sake of contradiction, assume that

$$\Pr_{x\sim\mathcal{D}}[f(x) = g(x)] > 1 - \delta' + \epsilon(1-\delta') = (1-\delta')(1+\epsilon)$$

for some $f \in \mathcal{F}^m$. We want to show that this violates the pseudodensity assumption on S. Let us develop the LHS: When do the functions $f(x)$ and $g(x)$ agree? This can only occur in two ways:

1. x comes from S and $f(x) = 1$: Because V_S appends 1 in front and $g(b, x) = b$, $g(x)$ will also be 1.

2. x comes from U and $f(x) = 0$: Because V_U appends 0 in front and $g(b, x) = b$, $g(x)$ will also be 0.

Now we compute the success probability for each of these two events:

1. The probability that x comes from S is δ'. Hence, the probability that f returns 1 is

$$\delta' \cdot \Pr_{x \in S}[f(x) = 1].$$

2. The probability that x comes from U is $1 - \delta'$. Probability that f returns 0 is

$$(1 - \delta')(1 - \Pr_{x \in U}[f(x) = 1]).$$

Hence,

$$\Pr_{x \sim \mathcal{W}}[f(x) = g(x)] = \delta' \cdot \Pr_{x \in S}[f(x) = 1] + (1 - \delta')(1 - \Pr_{x \in U}[f(x) = 1]).$$

By plugging in the statement that we are assuming by contradiction, it follows that

$$\Pr_{x \sim \mathcal{W}}[f(x) = g(x)] = \delta' \Pr_{x \in S}[f(x) = 1] + (1 - \delta')(1 - \Pr_{x \in U}[f(x) = 1]) > 1 - \delta' + \epsilon(1 - \delta') = (1 - \delta')(1 + \epsilon).$$

By dividing by $(1 - \delta')$, we obtain

$$\frac{\delta'}{1 - \delta'} \cdot \Pr_{x \in U}[f(x) = 1] + \frac{1 - \delta'}{1 - \delta'} \cdot (1 - \Pr_{x \in S}[f(x) = 1]) > 1 + \epsilon.$$

Since by definition $\delta' := \delta/(1+\delta)$, it follows that $\delta = \delta'/(1-\delta')$. Therefore, the expression simplifies to

$$\delta \Pr_{x \in S}[f(x) = 1] + 1 - \Pr_{x \in U}[f(x) = 1] > 1 + \epsilon.$$

(Notice that δ' is chosen as $\delta/(1 + \delta)$ precisely so that we obtain this simplification.) Re-arranging sides,

$$\Pr_{x \in U}[f(x) = 1] < \delta \Pr_{x \in S}[f(x) = 1] - \epsilon.$$

This contradicts the assumption that S is (F^m, ϵ, δ)-pseudodense in $\mathcal{U}_\mathcal{X}$. □

Step 2: Apply IHCL. Since we have shown that g is weakly hard, we can now apply IHCL to g to get a hardcore measure (with optimal density 2δ).

Remark 6.13. In this proof, given that g is weakly hard with respect to distribution $\mathcal{W}$, rather than with respect to the uniform distribution $\mathcal{U}_V$, we need to use a slightly more generalized version of IHCL than the one we stated in Chapter 4. The statement is exactly the same one as IHCL, except that the weakly hardness assumption is with respect to an arbitrary distribution $\mathcal{W}$ on V rather than only $\mathcal{U}_V$. See, e.g., Theorem 6 in [Imp09]. Notice that this matches the version of IHCL that we stated in Chapter 5 (Theorem 5.17), where the weakly hardness assumption was also with respect to an arbitrary distribution.

Specifically, given our parameters above, we get:

Claim 6.14. *There exists a measure μ over V of density $2\delta' - 2\epsilon(1-\delta')$ with respect to the distribution $\mathcal{W}$, such that g is $(\mathcal{F}, \epsilon\delta'/4)$-hardcore.*

Proof. This follows direclty from applying IHCL (Theorem 4.9), where the weakly hardness assumption is with respect to $\mathcal{W}$ instead of $\mathcal{U}_V$. □

Step 3: Proof of the density of the model. Let μ correspond to the measure from Claim 6.14. Let μ_S denote the restriction of μ on S and likewise let μ_U denote the restriction of μ on U. By the definition of the distribution $\mathcal{W}$ and by the lower bound on the density of μ given by Claim 6.14 (which is given by IHCL), it follows that

Using the bound on the density of μ:

$$d(\mu) = \delta' d(\mu_S) + (1-\delta')d(\mu_U) \geq 2\delta' - 2(1-\delta')\epsilon. \tag{6.15}$$

We will now show that μ_U corresponds to the measure that we are looking for; i.e., that μ_U is a model for S (thus proving the DMT). Intuitively, splitting $d(\mu)$ into $\delta' d(\mu_S)$ and $(1-\delta')d(\mu_U)$ is a way of "recovering uniformity": namely, by the DMT statement, the pseudodensity of S is with respect to the *uniform* distribution on the domain. The density of μ_S is with respect to the uniform distribution on S, and the density of μ_U is with respect to the uniform distribution on U. Hence, this splitting of $d(\mu)$ "reverts" the non-uniformity introduced by the distribution $\mathcal{W}$, which allows us to prove that μ_U is a dense measure with respect to the *uniform distribution*. Indeed:

Claim 6.16. *μ_U is a $(\delta - O(\epsilon))$-dense measure on U with respect to the uniform distribution.*

Proof. The intuition behind this is as follows: in order for μ to be a hardcore set, it must be split approximately evenly between U and S (up to an ϵ slack); otherwise, we could have an advantage by predicting the constant 0 or 1 value. That is, by definition of a hardcore set:

$$\left| \Pr_{(b,x)\sim D_\mu}[g(b,x) = 0] - 1/2 \right| \leq \epsilon\delta'/4,$$

$$\left| \Pr_{(b,x)\sim D_\mu}[g(b,x) = 1] - 1/2 \right| \leq \epsilon\delta'/4,$$

where $b \in \{0,1\}$. This is because the constant functions $\mathbf{0}$ and $\mathbf{1}$ are in $\mathcal{F}$ (Remark 2.2). Hence, the $1/2$ term in the two expressions above corresponds to how well we would do if we were predicting with the constant function. Hence, by adding together the two expressions together, we obtain that

$$\left| \Pr_{(b,x)\sim D_\mu}[g(b,x) = 1] - \Pr_{(b,x)\sim D_\mu}[g(b,x) = 0] \right| \leq \epsilon\delta'/2.$$

Hence,

$$\left| \delta' d(\mu_S) - (1-\delta')d(\mu_U) \right| \leq \epsilon\delta' d(\mu)/2 \leq \epsilon\delta'/2. \tag{6.17}$$

Solving for Equations 6.15 and 6.17, we obtain that

$$d(\mu_S) \geq 1 - O(\epsilon/\delta),$$

$$d(\mu_U) \geq \delta - O(\epsilon).$$

Therefore, the measure μ_U has density at least $\delta - O(\epsilon)$, as we wanted to show. □

Lastly, to conclude the proof of the DMT (Theorem 6.10), we need to show that μ_U is a good model for S.

Step 4: Proof of the indistinguishabilty of the model. Formally, we want to show:

Claim 6.18. *The measure μ_U is an $O(\mathcal{F}, \epsilon/\delta)$-model for S.*

By definition of a model, this corresponds to showing that D_S and D_{μ_U} are $(\mathcal{F}, \epsilon/\delta)$-indistinguishable.

Proof. First, recall that $d(\mu_S) \geq 1 - O(\epsilon/\delta)$. We claim that this implies that D_{μ_S} and the uniform distribution on S are statistically close: namely, their statistical distance is at most $O(\epsilon/\delta)$. This is because we can write D_{μ_S} as a convex combination of the uniform distributions on sets A of size $d(\mu_{\mathcal{X}})|S|$. Since each such distribution has statistical distance

$$2(1 - |A|/|S|) = 2(1 - d(\mu_S)) = O(\epsilon/\delta),$$

the same is true for D_{μ_S} [Imp09]. Therefore, for each $f \in \mathcal{F}$,

$$\left| \Pr_{x \sim S}[f(x) = 1] - \Pr_{x \sim D_{\mu_U}}[f(x) = 1] \right| \leq O(\epsilon/\delta) + \left| \Pr_{x \sim D_{\mu_S}}[f(x) = 1] - \Pr_{x \sim D_{\mu_U}}[f(x) = 1] \right|.$$

Now we need to show that the RHS is small, to conclude that the distinguishers $f \in \mathcal{F}$ cannot know distinguish whether x has been sampled from D_{μ_S} or from D_{μ_U}, which corresponds exactly to the notion of indistinguishability. Since g is $(\mathcal{F}, \epsilon\delta'/4)$-hardcore by Claim 6.14, it follows that

$$\frac{1}{2} + \epsilon\delta'/4 \geq \Pr_{x \sim D_\mu}[g(b,x) = f(b,x)] = \Pr_{x \sim D_\mu}[x \in V_S] \cdot \Pr_{x \sim D_{\mu_S}}[f(x) = 1] + \Pr_{x \sim D_\mu}[x \in V_U] \cdot \Pr_{x \sim D_{\mu_U}}[f(x) = 0].$$

Again by using the distinguishers $\mathbf{0}$ and $\mathbf{1}$, it must be that

$$\frac{1}{2} - \frac{\epsilon\delta'}{4} \leq \Pr_{x \sim D_\mu}[x \in V_S], \Pr_{x \sim D_\mu}[x \in V_U] \leq \frac{1}{2} + \frac{\epsilon\delta'}{4}. \tag{6.19}$$

Then, by combining Equations 6.15 and 6.19 it follows that

$$\Pr_{x \sim D_\mu}[x \in V_S] \cdot \Pr_{x \sim D_{\mu_S}}[f(x) = 1] - \Pr_{x \sim D_\mu}[x \in V_U] \cdot \Pr_{x \sim D_{\mu_U}}[f(x) = 1] \leq \frac{\epsilon\delta'}{2}. \tag{6.20}$$

Lastly, we need handle the signs separately. If $\Pr_{x \sim D_\mu}[x \in V_S] \leq \Pr_{x \sim D_\mu}[x \in V_U]$, then by Equation 6.19 it follows that

$$\Pr_{x \sim D_\mu}[x \in V_U] - \frac{\epsilon\delta'}{2} \leq \Pr_{x \sim D_\mu}[x \in V_S].$$

If we plug this into Equation 6.20, we conclude that

$$\left| \Pr_{x \sim D_{\mu_S}}[f(x) = 1] - \Pr_{x \sim D_{\mu_U}}[f(x) = 1] \right| \leq \frac{\epsilon\delta'}{1/2 - \epsilon\delta'/4}. \tag{6.21}$$

On the other hand, if $\Pr_{x \sim D_\mu}[x \in V_S] > \Pr_{x \sim D_\mu}[x \in V_U]$, the same argument holds symmetrically by using the distinguisher $-f$ instead of f. By Definition 6.2, Equation 6.21 corresponds exactly to stating that distributions D_{μ_S} and D_{μ_U} are $(\mathcal{F}, O(\epsilon/\delta)$-indistinguishable. Then, by the definition of a model (Definition 6.7), this corresponds to stating that μ_U is an $(\mathcal{F}, O(\epsilon/\delta))$-model for S, exactly as we wanted to show. □

Then, by Claim 6.16, we know that μ_U is a $(\delta, O(\epsilon))$-dense measure on U, and by Claim 6.18 we know that μ_U is an $(\mathcal{F}, O(\epsilon/\delta))$-model for S. Therefore, we have found the required model for S as specified in Theorem 6.10, thus proving the Dense Model Theorem.

6.3 Our proposed DMT++

After having understood the how IHCL implies DMT, we now propose our DMT++ statement, following the same general ideas outlined in Chapters 4 and 5 (i.e., by using a multicalibrated partition $\mathcal{P}$). One of the issues that complicate the generalization of DMT++ is that it is unclear how to generalize the notion of pseudodensity. That is, in IHCL++ we are able to remove the δ-weakly hardness assumption on g, and we generalized the density parameter δ by using the balance parameter k_p of g on each $P \in \mathcal{P}$. This balance parameter applies to an *arbitrary* function g. Likewise, for the DMT++ statement we hope to generalize the notion of (ϵ, δ)-pseudodensity, so that it can apply to an arbitrary set $S \subseteq \mathcal{X}$.

We are able to do this generalization by using the MC definition directly, instead of going through IHCL. That is, from the work of Trevisan et al. [TTV09], we know that a multiaccurate predictor can shown both IHCL and DMT separately. Hence, we will obtain DMT++ by starting with a multicalibrated predictor instead, rather than through the route MC → IHCL++ → DMT++, which is the route we took to obtain PAME++ in Chapter 5. However, we use several ideas from the proof that IHCL implies the DMT, which is why we included the proof in Section 6.2: in particular, the instantiation of the distinguishers, the augmentation of the domain, and the function g.

Our DMT++ and subsequent corollary are as follows:

Theorem 6.22 (DMT++)**.** *Let $\mathcal{X}$ be a finite domain, let $S \subseteq \mathcal{X}$, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, and let $\epsilon, \gamma > 0$. Let $U = \mathcal{X} \setminus S$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$, which satisfies that for each $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, distributions $\mathcal{U}_{P \cap S}$ and $\mathcal{U}_{P \cap U}$ are $(\mathcal{F}, \epsilon_p)$-indistinguishable for all $P \in \mathcal{P}$, where $\epsilon_p = \epsilon \cdot v_p \cdot (1 - v_p)$ for*

$$v_p = \frac{|P \cap S|/|S|}{2\eta_p}, \quad \eta_p = \frac{|P \cap S|}{2|S|} + \frac{|P \cap U|}{2|U|}.$$

That is, $\mathcal{U}_{P \cap U}$ is an $(\mathcal{F}, \epsilon_p)$-model for the corresponding set $P \cap S$ with density $|P \cap U|/|U|$.

Definition 6.23. *For notation purposes, we will use the abbreviations* $S_p := P \cap S$ *and* $U_p := P \cap U$ *for each* $P \in \mathcal{P}$.

Before we proceed to the proof of DMT++, we unpack what the statement is saying. We begin with two sets $S, U := \mathcal{X} \setminus S$, and we are looking for a partition of the domain:

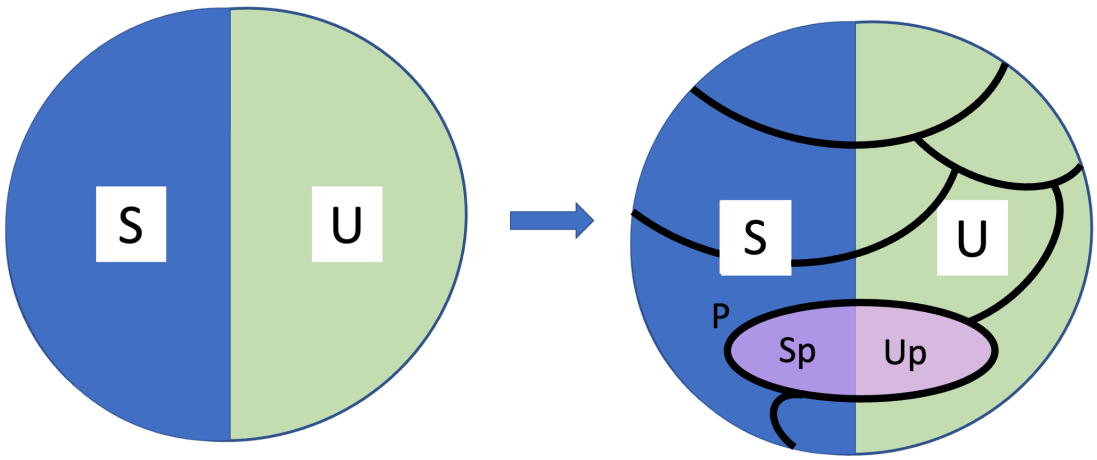

Figure 6.2: Visual depiction of the DMT++ statement (Theorem 6.22).

The picture highlights an example set $P \in \mathcal{P}$. For each $P \in \mathcal{P}$, the set P splits into the two parts S_p and U_p, where S_p corresponds to the subset of P that is contained in S and U_p corresponds to the subset of P that is contained in U. (Notice that either S_p or U_p can be empty, which occurs in the cases where P is entirely contained in S or entirely contained in U.) Then, the DMT++ statement requires the uniform distribution over U_p to be $(\mathcal{F}, \epsilon_p)$-indistinguishable from the uniform distribution over S_p, where ϵ_p depends on the relative sizes of S_p and U_p.

As in the case of ICHL++ and PAME++, when we compare DMT to the original DMT++, we see that DMT++ is able to replicate the DMT locally within each $P \in \mathcal{P}$; namely, there is a model for a subset of S within each set $P \in \mathcal{P}$. Moreover, DMT++ removes the assumption of DMT: in the case of DMT, the set $S \subseteq \mathcal{X}$ is assumed to be (ϵ, δ)-pseudodense. In the case of DMT++, the result holds for an arbitrary set $S \subseteq \mathcal{X}$, without any pseudodensity assumption. The same is true for IHCL++ and PAME++, where we are able to replicate the original theorems locally while removing the assumption. We further summarize the similarities between the three theorems in Chapter 7.

Intuition behind the proof. The idea behind the proof of Theorem 6.22 is the following. We will show that a multicalibrated partition $\mathcal{P}$ satisfies the conditions stated in the DMT++ theorem. But in order to apply the MC partition theorem (Theorem 2.29), we need invoke it with a family of distinguishers $\mathcal{F}$ and a function g. To do so, we will follow Impagliazzo's idea and define $\mathcal{F}$ and g similar to how we did it in Section 6.2 when showing that IHCL implies DMT. Namely, we will augment the domain from $\mathcal{X}$ to V and consider the class of distinguishers that ignore the first bit of $x \in V$. Similarly, we will define the function g as $g(b, x) = b$. Then, we will see that this definition of g allows us to relate the parameter $v_p := \mathbb{E}_P[g(x)]$ for each $P \in \mathcal{P}$ to the "skewness" of each set $P \in \mathcal{P}$. That is, how contained P is into either S or U. In particular, we will see the following:

- The larger v_p, the more it is contained in S.
- The smaller v_p, the more it is contained in U.

This is because $g(b, x) = 1$ if and only if $x \in S$, by definition of g. Since g is boolean, this means that $v_p \cdot |P|$ equals exactly the number of $x \in P$ such that $g(x) = 1$.

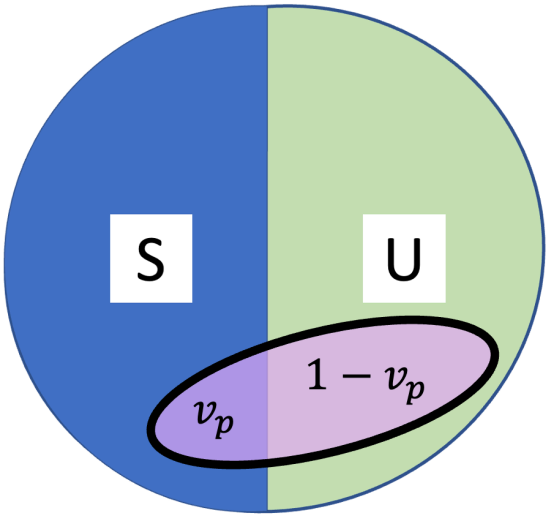

Figure 6.3: Using IHCL to prove the DMT. The parameter v_p determines the skewness of the set P into S.

As in the case of IHCL++, our DMT++ theorem is stronger and more general than the original DMT theorem for the following reasons:

- In Theorem 6.22, we remove the (ϵ, δ)-pseudodensity assumption on the set $S \subseteq \mathcal{X}$, but still obtain the indistinguishability condition; namely, we find a distribution within each P such that the set $P \cap S$ is $(\mathcal{F}, \epsilon)$-indistinguishable from it. The caveat is that this indistinguishability parameter ϵ_p depends on the size of P and the ratio $|P \cap S|/|S|$, which can make the indistinguishability parameter large.
- We provide a generalization of the notion of pseudodensity, and we are able to argue about the density of each model with the general parameter $|P \cap U|/|U|$, rather than using the parameter δ in the original DMT statement (which we do not have, since we do not have the (ϵ, δ)-pseudodensity assumption on S). When we show that our DMT++ implies the original DMT statement, as it was the case in IHCL++ and PAME++, we will see that the density guarantees on each distribution $\mathcal{U}_{P \cap U}$ imply that, when we "glue" the sets $P \in \mathcal{P}$ together and bring back the assumption that S is (ϵ, δ)-pseudodense, we obtain that the "global" model for S has density $O(\delta - O(\epsilon))$, as in the original DMT (Theorem 6.10).
- In our ++ theorem, the original DMT occurs both "locally" (on each $P \in \mathcal{P}$) and "globally" (on $\mathcal{X}$). Theorem 4.11 states that IHCL occurs locally; namely, we obtain a hardcore distribution $\mathcal{H}_P$ within each $P \in \mathcal{P}$. However, we can always "glue" the different hardcore measures together $\mathcal{H}_p$ in order to obtain a hardcore measure $\mathcal{H}$ on $\mathcal{X}$. Since g is strongly hard on each $\mathcal{H}_p$, g will also be strongly hard on the "glued" hardcore measure $\mathcal{H}$. In Section 4.2,

we will show that if we glue the different hardcore measures $\mathcal{H}_p$ together weighted by their corresponding size parameter $\eta_p := |P|/|\mathcal{X}|$, and if we bring back the assumption that g is δ-weakly hard (which is the key assumption in the original IHCL statement), then the glued hardcore set $\mathcal{H}$ has density at least 2δ on $\mathcal{U}_{\mathcal{X}}$. That is, we have recovered the original IHCL statement from our IHCL++ theorem.

Proof of Theorem 6.22. We begin by augmenting the domain $\mathcal{X}$ as we did in Section 6.2. From $\mathcal{X}$, we build the sets

$$V_S = \{(1, x) \mid x \in S\}$$
$$V_U = \{(0, x) \mid x \in \mathcal{X} \setminus S\}.$$

We then construct the augmented domain V as

$$V = V_S \cup V_U.$$

Given the class of distinguishers $\mathcal{F}$ on $\mathcal{X}$, the class of distinguishers on V corresponds to the same set of functions $f \in \mathcal{F}$, except that they ignore the first bit $b \in \{0, 1\}$ of the elements $x \in V$. We define the following function $g : V \to \{0, 1\}$:

$$g(b, x) = b.$$

That is g, indicates whether x "comes from" S or U.

We can now invoke the approximate MC partition theorem (Theorem 2.29) in the domain V with $\mathcal{F}, g, \epsilon, \gamma$, and where $\mathcal{D}$ corresponds to the distribution $\frac{1}{2}\mathcal{U}_{V_S} + \frac{1}{2}\mathcal{U}_{V_U}$ over V. (Defining $\mathcal{D}$ in this way corresponds to the same way in which we defined the distribution in the proof that IHCL implies DMT.) This gives us a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ satisfying

$$\left| \mathop{\mathbb{E}}_{x \sim P}[f(x) \cdot (g(x) - v_p)] \right| \leq \epsilon$$

for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$. We remark that we can write $x \sim P$ instead of $\mathcal{P}(\mathcal{D})|_P$ because in this case $\mathcal{D}$ corresponds to the uniform distribution over $\mathcal{X}$.

We claim that this partition $\mathcal{P}$ satisfies the conditions of Theorem 6.22. Fix any $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$. First, we notice that since g is boolean and $g(x) = 1$ if and only if $x \in S$, by the definition of $\mathcal{D}$ as $\frac{1}{2}\mathcal{U}_{V_S} + \frac{1}{2}\mathcal{U}_{V_U}$, for each such $P \in \mathcal{P}$,

$$v_p = \frac{|P \cap S|/|S|}{|P \cap S|/|S| + |P \cap U|/|U|} = \frac{|P \cap S|/|S|}{2\eta_p}.$$

Recall that S_p denotes the set $P \cap S$ and U_p denotes the set $P \cap S$. We need to show that the distributions $\mathcal{U}_{S_p}$ and $\mathcal{U}_{U_p}$ are $(\mathcal{F}, \epsilon_p)$-indistinguishable, where

$$\epsilon_p = \epsilon \cdot v_p \cdot (1 - v_p).$$

By definition of indistinguishability, this corresponds to showing that

$$\left| \mathop{\mathbb{E}}_{x\sim S_p}[f(x)] - \mathop{\mathbb{E}}_{x\sim U_p}[f(x)] \right| \leq \epsilon \cdot v_p \cdot (1 - v_p).$$

(Notice that we write $\mathbb{E}$ instead of the formulation with Pr that appears in Definition 2.1. However, since the functions $f \in \mathcal{F}$ are boolean, the two notions are equivalent by the fundamental bridge.)

Indeed, we see that

$$\left| \mathop{\mathbb{E}}_{x\sim S_p}[f(x)] - \mathop{\mathbb{E}}_{x\sim U_p}[f(x)] \right| = \left| \frac{\mathbb{E}_{x\sim P}[f(x)g(x)]}{v_p} - \frac{\mathbb{E}_{x\sim P}[f(x)(1-g(x))]}{1-v_p} \right| \leq \frac{\epsilon}{v_p \cdot (1-v_p)},$$

as needed. □

Remark 6.24. We remark that the indistinguishability parameter ϵ_p for each P, as it was the case for IHCL++ and PAME++, degrades with the balance parameter of the set k_p. Namely, recall that $k_p = \min\{v_p, 1-v_p\}$ denotes the balance of g on P. Then, $\frac{1}{2} \cdot k_p \leq v_p \cdot (1-v_p) \leq k_p$, and hence k_p and $v_p \cdot (1-v_p)$ are equivalent up to a factor of 2. Therefore, we think of the term $v_p \cdot (1-v_p)$ that appears in the denominator of ϵ_p as also corresponding to the balance of g on the set $P \in \mathcal{P}$.

6.3.1 Recovering the original DMT from DMT++

Having proved the DMT++ theorem, we now show how to recover the original DMT theorem. As we did for IHCL and PAME, the key idea is to "glue together" the models $\mathcal{U}_{P\cap U}$ for each "good" $P \in \mathcal{P}$. Namely, those $P \in \mathcal{P}$ that have $\eta_p \geq \gamma$ and $k_p \geq \tau$ for parameters γ, τ (Definition 4.13 from Chapter 4).

Recall that in the DMT statement, we are bringing back the assumption that the set S is $(\mathcal{F}^m, \epsilon, \delta)$-pseudodense. As we did in Chapter 4, we begin by showing an intermediate proposition:

Proposition 6.25. *Let $\mathcal{P}$ be a partition of $\mathcal{X}$ as in Theorem 6.22. Moreover, assume that S is (F^m, ϵ, δ)-pseudodense for some $\delta > 0$. Then, for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$,*

$$\delta \cdot \frac{|S_p|}{|S|} \leq \frac{|U_p|}{|U|} + \epsilon.$$

Proof. Recall that by definition of pseudodensity, we know that for all $f \in \mathcal{F}^m$,

$$\delta \cdot \Pr_{x\sim S}[f(x) = 1] - \epsilon \leq \Pr_{x\sim U}[f(x) = 1].$$

Fix some $P \in \mathcal{P}$. We proceed similar to the proof of Proposition 4.12 in Chapter 4. Let $f_m \in \mathcal{F}_{t,q}$ where $t = O(1/(\epsilon'^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon'))$, $q = O(1/\epsilon'^2)$ be the partition membership function for $\mathcal{P}$ as given by Definition 2.28. That is, $P_i = f^{-1}(i)$ for all of the k sets $P_i \in \mathcal{P}$. Using this f_m, we construct the following function $f\colon \mathcal{X} \to \{0,1\}$:

$$f(x) := \begin{cases} 1 & \text{if } x \in P, \\ 0 & \text{otherwise .} \end{cases}$$

That is, we can think of f as the indicator function for P. We claim that $f \in \mathcal{F}_{t,q}$ for the same parameters t, q. Indeed, let C_m be the oracle-aided circuit that computes f_m. It is enough that we hard-wire the values 0, 1 as described above. (To know whether it should be 0 or 1 for each $x \in \mathcal{X}$, we use a look-up table.) Hence, the circuit that computes f is of size $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon)) + |\mathcal{P}|$ and continues to have $q = O(1/\epsilon^2)$ oracle gates (the same as for f_m) [Bar22, §9.1.1.]. Since $|\mathcal{P}| = O(1/\epsilon)$ by Theorem 4.11, it follows that $f \in \mathcal{F}_{t,q}$, since the term $O(1/\epsilon)$ is absorbed into t.

Then, by applying the pseudodensity assumption to this function f, we obtain that

$$\delta \cdot \Pr_{x \sim S}[f(x) = 1] - \epsilon \leq \Pr_{x \sim U}[f(x) = 1].$$

Then, since

$$\Pr_{x \sim S}[f(x) = 1] = \frac{|S_p|}{|S|}, \quad \Pr_{x \sim U}[f(x) = 1] = \frac{|U_p|}{|U|},$$

it follows that

$$\delta \cdot \frac{|S_p|}{|S|} \leq \frac{|U_p|}{|U|} + \epsilon.$$

Re-arranging, we obtain the inequality stated in Proposition 6.25. □

Given this proposition, we can now proceed to the proof of DMT

Proof of DMT using DMT++. Let $\mathcal{F}, \mathcal{X}, \epsilon, \delta$ be the assumption parameters in the DMT statement (Theorem 6.10). We define the parameters $\epsilon' := \epsilon^2\delta$, $\gamma := \epsilon\epsilon'$, and invoke the DMT++ theorem (Theorem 6.22) with these parameters ϵ', γ. By DMT++ (Theorem 6.22), we obtain a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon'^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon'))$, $q = O(1/\epsilon'^2)$, $k = O(1/\epsilon')$ such that, for each $P \in \mathcal{P}$ where $\eta_p \geq \gamma = \alpha\epsilon'$, there exists an $(\mathcal{F}, \epsilon'_p)$-model $\mathcal{U}_{P \cap U}$ for the corresponding set $P \cap S$. Given these "little" measures $\mathcal{U}_{P \cap U}$, we construct the claimed measure μ for S as follows: we define μ to be the "weighted average" of the models $\mathcal{U}_{P \cap U}$ such that $P \in \mathcal{P}$ is (γ, τ)-good. Formally, for each $x \in U$,

$$\mu(x) = \frac{|S_p|}{|S|} \cdot \frac{1}{|U_p|} \cdot \mathbb{1}_G(P),$$

where P corresponds to the unique $P \in \mathcal{P}$ such that $x \in P$ (which is unique since $\mathcal{P}$ is a partition). That is, for the "good" P, μ assigns a probability mass of $(|S_p|/|S|) \cdot (1/|U_p|)$, where the sets S_p, U_p for each $P \in \mathcal{P}$ are given by Theorem 6.22. Intuitively, μ corresponds to the weighted average of the U_p obtained from DMT++. The expression $(|S_p|/|S|) \cdot (1/|U_p|)$ should be understood as follows: in order to choose a $x \in U$, we first choose a set P with probability $|S_p|/|S|$, and then choose one of the points in $U_p \subseteq U$ for that P uniformly over U_p. Note also that μ only assigns non-zero mass to points in U, given that we are building a model for S.

Let v_p as in Theorem 6.22. First, by Proposition 6.25 we know that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$,

$$\frac{|S_p|}{|S|} \cdot \frac{1}{|U_p|} \leq \left(\frac{1}{\delta} + \frac{\epsilon' \cdot |U|}{\delta \cdot |U_p|}\right) \cdot \frac{1}{|U|}.$$

Then, since

$$\frac{|U_p|}{|U|} = 2\eta_p(1 - v_p),$$

it follows that

$$\frac{1}{\delta} + \frac{\epsilon' \cdot |U|}{\delta \cdot |U_p|} = \frac{1}{\delta} + \frac{\epsilon'}{2\delta\eta_p(1-v_p)} = \frac{1}{\delta \cdot (1 - O(\epsilon'/(\eta_p(1-v_p))))}.$$

Recall that we only "glue up" the parts P that are (γ, τ)-good, and recall that $\mathbb{1}_G(P)$ returns 1 if $\eta_p \geq \gamma$ and $k_p \geq \tau$, and 0 otherwise. Then, following a similar analysis as the one we did when recovering the original IHCL from IHCL++ in Chapter 4, it follows that

$$\mathop{\mathbb{E}}_{P \sim \mathcal{P}} \left[\frac{\epsilon'}{v_p \cdot (1 - v_p)} \cdot \mathbb{1}_G(P) \right] \leq \frac{\epsilon'}{\tau}.$$

Therefore, by plugging in the definitions of the parameters, namely $\epsilon' = \epsilon^2\delta$ and $\tau = \epsilon\delta$, the density of the "glued up" measure μ (when we only glue up the (γ, τ)-good sets) is equal to $\delta \cdot (1 - O(\epsilon))$.

Next, we show the indistinguishability condition; namely, that μ is an $\mathcal{F}, O(\epsilon/\delta)$-model for S. First we assume that all $P \in \mathcal{P}$ are such that $\eta_p \geq \gamma$. By Theorem 6.22, we know that for all $P \in \mathcal{P}$,

$$\left| \Pr_{x \sim S_p} f[(x) = 1] - \Pr_{x \sim U_p}[f(x) = 1] \right| \leq \epsilon'_p.$$

By the law of total probabilty, it follows that

$$\Pr_{x \in S}[f(x) = 1] = \sum_{P \in \mathcal{P}} \Pr[f(x) = 1 \mid x \in S_p] \cdot \Pr[x \in S_p] = \sum_{P \in \mathcal{P}} \Pr_{x \sim S_p}[f(x) = 1] \cdot \frac{|S_p|}{|S|}.$$

Next, we study the quantity $\Pr_{x \sim \mu}[f(x) = 1]$. By definition of μ and by the law of total probability, it follows that

$$\Pr_{x \sim \mu}[f(x) = 1] = \sum_{P \in \mathcal{P}} \Pr[f(x) = 1 \mid x \in U_p] \cdot \Pr[x \in U_p] = \sum_{P \in \mathcal{P}} Pr_{x \sim U_p}[f(x) = 1] \cdot \frac{|S_p|}{|S|},$$

since by the definition of μ, all values $x \sim \mu$ are in U, and hence $\mu(x)$ is never contained in S. Therefore, by the triangle inequality and the expressions above, it follows that

$$\left| \Pr_{x \sim S}[f(x) = 1] - \Pr_{x \sim \mu}[f(x) = 1] \right| = \left| \sum_{P \in \mathcal{P}} \Pr_{x \sim S_p}[f(x) = 1] \cdot \frac{|S_p|}{|S|} - \sum_{P \in \mathcal{P}} \Pr_{x \sim U_p}[f(x) = 1] \cdot \frac{|U_p|}{|U|} \right|$$

$$= \left| \sum_{P \in \mathcal{P}} \left(\Pr_{x \sim S_p}[f(x) = 1] \cdot \frac{|S_p|}{|S|} - \Pr_{x \sim U_p}[f(x) = 1] \cdot \frac{|U_p|}{|U|} \right) \right|$$

$$\leq \frac{|S_1|}{|S|} \cdot \left| \Pr_{x \sim S_1}[f(x) = 1] - \Pr_{x \sim U_1}[f(x) = 1] \right| + \cdots + \frac{|S_{|\mathcal{P}|}|}{|S|} \cdot \left| \Pr_{x \sim S_{|\mathcal{P}|}}[f(x) = 1] - \Pr_{x \sim U_{|\mathcal{P}|}}[f(x) = 1] \right|.$$

By applying Theorem 6.22 to all of the summands, the above expression becomes

$$\frac{|S_1|}{|S|} \cdot \epsilon'_1 + \cdots + \frac{|S_{|\mathcal{P}|}|}{|S|} \cdot \epsilon'_{|\mathcal{P}|} = \sum_{P \in \mathcal{P}} \epsilon'_p \cdot \frac{|S_p|}{|S|} = \sum_{P \in \mathcal{P}} \frac{\epsilon'}{v_p \cdot (1 - v_p)} \cdot \frac{|S_p|}{|S|}$$

$$\leq \sum_{P \in \mathcal{P}} \frac{\epsilon'}{v_p \cdot (1 - v_p)} \cdot \frac{1}{\delta} \cdot \frac{|U_p|}{|U|} = \frac{1}{\delta} \sum_{P \in \mathcal{P}} \frac{\epsilon'}{v_p \cdot (1 - v_p)} \cdot \frac{|U_p|}{|U|}.$$

Since $|U_p|/|U| = 2\eta_p \cdot (1 - v_p)$, it follows that each term in the summand is at most $2\tau/v_p \leq 2\epsilon'/\tau$. Therefore, by a similar analysis as in the case of the density parameter,

$$\mathop{\mathbb{E}}_{P \sim \mathcal{P}} \left[\frac{\epsilon'}{\tau} \cdot \mathbb{1}_G(P) \right] \leq \epsilon,$$

since recall that we only glue up the pieces P that are (γ, τ)-good. By plugging in the definitions of the parameters, namely $\epsilon' = \epsilon^2\delta$ and $\tau = \epsilon\delta$, we obtain that $\epsilon'/\tau = \epsilon$. Therefore, by plugging everything together, we obtain that μ is an $(\mathcal{F}, O(\epsilon/\delta)$-model for S. As we did in Chapter 4, we can modify μ so that we obtain density exactly $\delta - O(\epsilon)$ while maintaining $O(\epsilon/\delta)$-indistinguishability for the model. Thus, we have recovered the original DMT theorem.

□

7

The General Picture

Now what is science? (...) It is before all a classification, a manner of bringing together facts which appearances separate, though they are bound together by some natural and hidden kinship.

Henri Poincaré, The Value of Science (1907)

We conclude Part II of this thesis by describing the underlying principle behind the IHCL++ (Chapter 4), PAME++ (Chapter 5), and DMT++ (Chapter 6) statements, and how our generalizations of the original IHCL, PAME, and DMT theorems relate to each other. Let us first re-state the three ++ theorems together:

Theorem 7.1 (IHCL++, measure version)**.** *Let $\mathcal{X}$ be a finite domain, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, let $g\colon \mathcal{X} \to \{0,1\}$ be an arbitrary function, and let $\epsilon, \gamma > 0$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ which satisfies that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, there exists a distribution $\mathcal{H}_P$ in P of density $2k_p$ in $\mathcal{U}_P$ such that g is $(\mathcal{F}, \epsilon/k_p)$-strongly hard on $\mathcal{H}_P$. That is,*

$$\forall f \in \mathcal{F}, \quad \Pr_{x \sim \mathcal{H}_P}[f(x) = g(x)] \leq 1/2 + \frac{\epsilon}{k_p}.$$

Theorem 7.2 (PAME++)**.** *Let $\mathcal{F}$ be any class of functions, let (X, B) be a joint distribution on $\{0,1\}^n \times \{0,1\}$ where $X = \mathcal{U}_n$ and $B = g(X)$ for an arbitrary boolean function g, and let $\epsilon, \gamma > 0$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$ which satisfies that for all $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, $B|_P$ has non-uniform $(\mathcal{F}, \epsilon/k_p)$-PAME at least $\log(1/(1-k_p)) = H_\infty(g(\mathcal{U}_p))$ given $X|_P$, where $B|_P$ denotes the restriction of B on P and $X|_P$ the restriction of X on P.*

Theorem 7.3 (DMT++)**.** *Let $\mathcal{X}$ be a finite domain, let $S \subseteq \mathcal{X}$, let $\mathcal{F}$ be a family of functions $f\colon \mathcal{X} \to \{0,1\}$, and let $\epsilon, \gamma > 0$. Let $U = \mathcal{X} \setminus S$. There exists a partition $\mathcal{P} \in \mathcal{F}_{t,q,k}$ of $\mathcal{X}$ with $t = O(1/(\epsilon^4\gamma) \cdot \log(|\mathcal{X}|/\epsilon))$, $q = O(1/\epsilon^2)$, $k = O(1/\epsilon)$, which satisfies that for each $P \in \mathcal{P}$ such that $\eta_p \geq \gamma$, distributions $\mathcal{U}_{P\cap S}$ and $\mathcal{U}_{P\cap U}$ are $(\mathcal{F}, \epsilon_p)$-indistinguishable for all $P \in \mathcal{P}$, where*

$$\epsilon_p = \epsilon \cdot \left(\frac{|P \cap S|/|S|}{2\eta_p}\right) \cdot \left(1 - \frac{|P \cap S|/|S|}{2\eta_p}\right).$$

That is, $\mathcal{U}_{P\cap U}$ is an $(\mathcal{F}, \epsilon_p)$-model for the corresponding set $P \cap S$ with density $|P \cap U|/|U|$.

Each of the three theorems deals with a different **object**:

- In the case of IHCL, we are dealing with a *function* g.
- In the case of PAME, we are dealing with a joint *distribution* (X, B).
- In the case of DMT, we are dealing with a *set* S.

In the case of PAME, in this thesis we restricted the PAME theorem to the case where $B = g(X)$ for some function g, but more generally the PAME theorem deals with an arbitrary joint distribution (X, B), and we could generalize our PAME++ theorem accordingly.

In the original theorems, we have a certain **assumption** that we must require for each object in order for the theorem to be true:

- In the case of IHCL, function g is assumed to be δ-*weakly hard* for some parameter δ.
- In the case of PAME, distribution B is assumed to be $(1 - 2^{-r})$-*hard to predict* for some parameter r.
- In the case of DMT, set S is assumed to be (ϵ, δ)-*pseudodense* for some parameters ϵ, δ.

Given these assumptions, the original IHCL, PAME, and DMT statements are able to **conclude** the following:

- In the case of IHCL, we find a 2δ-dense hardcore set/measure.
- In the case of PAME, we find that B has PAME at least r.
- In the case of DMT, we find a δ-dense model for S.

Importantly, there are two separate parts in these conclusions. One deals with a sort of **density** guarantee (which is captured by the δ parameter) and another deals with an **indistinguishability** guarantee (which is captured with the parameter ϵ):

- In the case of IHCL, the density guarantee ensures that the hardcore sets/measures have density at least 2δ on $|\mathcal{X}|$ (or on $\mathcal{U}_\mathcal{X}$ for the measure version). The indistinguishability guarantee ensures that g is indeed ϵ-strongly hard on the set/measure (hence making it a "hardcore" set/measure).

- In the case of PAME, to show that B has PAME at least r we show that there exists a random variable C jointly distributed with X such that 1) C has average min-entropy at least r, and 2) (X, C) and (X, B) are ϵ-indistinguishable. Hence, we can understand condition 1) as a type of "density" guarantee, and condition 2) corresponds to an indistinguishability guarantee.

- In the case of DMT, the density guarantee ensures that the model has density at least δ, and the indstinguishability guarantee ensures that the model is indeed a model; namely, that the model is indstinguishable from the distribution induced by the set S.

One of the key ideas behind our ++ generalizations is that we should keep a separation between the density guarantees and the indistinguishability guarantees. This is precisely what allows us to remove the assumption from the original theorems in our ++ generalizations while maintaining the same type of conclusion; namely, a multicalibrated partition is what allows us to prove the indistinguishability guarantee in the conclusions, without having to rely on the assumptions that are necessary in the original theorems. Because we remove the assumptions from the original theorems, we no longer have parameter δ (which we think of as the "density") parameter in the ++ theorem. Thus, the ++ theorems require generalizing the δ "density" parameter so that it can apply to an arbitrary object. In our ++ theorems, this δ parameter is generalized by using the balance parameter k_p and the size parameter η_p of the set P in the partition instead.

With this in mind, now summarize the key underlying principles behind our ++ theorems. First, all of our ++ theorems consist of providing a partition $\mathcal{P}$ of the domain such that the original theorem "occurs" locally on each piece P of the partition $\mathcal{P}$. Because we prove our ++ theorems using a multicalibrated partition $\mathcal{P}$, one of the key aspect of our proofs is showing that a multicalibrated partition is such that the original theorems are occurring "locally" on each $P \in \mathcal{P}$. In particular, the indistinguishability parameter ϵ then depends on some parameter of the piece P (either on k_p, the balance of g on P, or on η_p, the relative size of P in the domain).

All of our ++ theorems remove the assumptions on the objects that are necessary in the original theorems:

- In the case of IHCL++, g is an arbitrary function.

- In the case of PAME++, B is an arbitrary distribution.

- In the case of DMT++, S is an arbitrary model.

The following describes the type of indistinguishability guarantees achieved by our ++ theorems:

- In the case of IHCL++, we find a hardcore set within each piece P of the partition. This corresponds to the indstinguishability guarantee, given that g is strongly hard on the hardcore set. In particular, this strong hardness is with respect to the indistinguishability parameter ϵ/k_p, which varies on each $P \in \mathcal{P}$.

- In the case of PAME++, we find a random variable $C|_P$ defined on each P such that $(X|_P, B|_P)$ and $(X|_P, C_P)$ are ϵ/k_p-indistinguishable. Again, the indistinguishability parameter on each $P \in \mathcal{P}$ depends on the parameter k_p, which varies on each $P \in \mathcal{P}$. Namely, the more balanced g is on P, the better the indistinguishability guarantee becomes.

- In the case of DMT++, we find a distribution defined on each $P \in \mathcal{P}$ such that it is a model for a corresponding set inside P. The indistinguishability parameter for the model, ϵ_p, also varies in each $P \in \mathcal{P}$.

We understand each of these ϵ_p as a "local" indistinguishability guarantee of the object on each $P \in \mathcal{P}$.

Regarding the density guarantees, as we have explained, each of the ++ theorems required generalizing the density parameters, which we do not have in the ++ theorems given that we removed the assumptions. In each ++ theorem, achieving a generalized density guarantee is as follows (it is thus helpful to think of k_p as a generalization of the parameter δ):

- In the case of IHCL++, the density of each of the hardcore set that exists on every $P \in \mathcal{P}$ is given by $2k_p$.
- In the case of PAME++, each of the random variable $C|_P$ that exists on every $P \in \mathcal{P}$ has average min-entropy at least $\log(1/(1-k_p))$.
- In the case of DMT++, each of the models that exists on every $P \in \mathcal{P}$ has density $|P \cap U|/|U|$, where U corresponds to the complement of the set S in the domain.

Lastly, the other key aspect of our ++ theorems, and which is also a unifying principle behind our ++ theorems, is that by gluing the objects that we have found on each $P \in \mathcal{P}$ and by bringing back the assumption that is present in the original theorems, we then recover exactly the original theorem (both the density guarantee and the indistinguishability guarantee). We understand this as saying that our ++ theorems hold both "locally" and "globally": locally because we find the object of interest within each $P \in \mathcal{P}$, and globally because when we glue those "little" objects together, we obtain a large object over the entire domain that satisfies the guarantees that we are interested it (i.e., both density and indistinguishability). Namely:

- In the case of IHCL++, when we glue together the hardcore sets present in all $P \in \mathcal{P}$, we obtain a hardcore set over the domain, which recovers the original "global" indistinguishability parameter ϵ. Moreover, when we bring back the assumption that g is δ-weakly hard, our "local" density guarantee (namely, $2k_p$) implies that the glued hardcore set is 2δ-dense on the domain. Hence, these two conditions together (namely, the global indistinguishability guarantee and the global density guarantee) recover the original IHCL.
- In the case of PAME++, when we glue together the random variables $C|_P$ preset in all $P \in \mathcal{P}$, we obtain a random variable C over the domain. We show that this C is such that (X, C) and (X, B) are ϵ-indistinguishable, thus recovering the original "global" indistinguishability guarantee. Moreover, when we bring back the assumption that B is $(1-2^{-r})$-hard to predict, which in our thesis corresponds to $B = g(X)$ where g is δ-weakly hard, then we obtain that the glued distribution C has average min-entropy at least r. Hence, these two conditions together (namely, the global indistinguishability guarantee and the global density guarantee) recover the original PAME theorem.

- In the case of DMT++, when we glue together the models present in all $P \in \mathcal{P}$, we obtain a distribution over the domain such that it is an ϵ-model for the set S. Hence, this recovers the original "global" indistinguishability guarantee. Moreover, when we bring back the assumption that S is (ϵ, δ)-pseudodense , then we obtained that the glued model has density δ. Hence, these two conditions together (namely, the global indistinguishability guarantee and the global density guarantee) recover the original DMT theorem.

Hence, all of these observations together explain the parallels between the IHCL++, PAME++, and DMT++ statements and describe their underlying principles.

Remark 7.4. In the discussion in this chapter, by simplicity, we always say that the guarantees hold for all $P \in \mathcal{P}$. As we have seen, we actually only consider the sets P such that $\eta_p \geq \gamma$ for some parameter γ, given that we are using the notion of approximate multicalibration.

8

Conclusions and Future Work

In this thesis, we have obtained stronger and more general versions of theorems that are implied by the notion of multiaccuracy (i.e., by the regularity lemma of Trevisan et al. [TTV09]) by considering the notion of multicalibration instead. In particular, we have obtained stronger and more general versions of Impagliazzo's Hardcore Lemma, the theorem of Vadhan and Zheng that characterizes the notion of pseudoentropy, and the Dense Model Theorem. The starting point of this thesis consisted of the observation that the regularity lemma of Trevisan et al. corresponds to the notion of multiaccuracy that has been recently developed in the field of algorithmic fairness. In algorithmic fairness, the notion of multicalibration has been proposed as a strenghtening of the notion of multiaccuracy, and the fairness literature has shown how to construct a multicalibrated predictor. Then, by using a multicalibrated predictor as our main tool, we cast it back into the realm of complexity theory and study how the implications of the regularity lemma are modified. In particular, since multicalibration is a stronger notion than multiaccuracy, in doing so we are able to obtain stronger and more general versions of the original theorems, all of which are fundamental theorems in theoretical computer science that have been known for years.

Our key observation for proving our strengthened and more general IHCL++, PAME++, and DMT++ theorems is that a multicalibration partition of the domain yields some sort of "indistinguishability for free" on each piece P of the partition. This allows us to replicate the original theorems within each piece P while maintaining the same indistinguishability conclusion as in the original theorem. We are also able to provide a density guarantee inside each P such that, when we glue the objects of interest that we have obtained in each partition and bring back the assumption of the original theorems, we recover the original density and indistinguishability parameters. Another key aspect of our proofs has been to exploit the connections that exist between the IHCL, PAME, and DMT theorems. Namely, after showing our new IHCL++ theorem, we obtain our PAME++ and DMT++ from the connections between IHCL and PAME and between IHCL and DMT. In particular, to prove our PAME++ statement, first we show that IHCL implies PAME, and then we plug our IHCL++ statement into this connection. In the case of DMT, while we prove DMT++ directly from a multicalibrated partition rather than plugging IHCL++ into the proof that IHCL implies DMT, in doing so we use some key ideas from the proof that IHCL implies DMT.

The main conclusion of our thesis is that, by casting back notions and techniques that have recently originated in algorithmic fairness back to the realm of complexity theory, we obtain a deep and fruitful connection between the two fields. We expect this connection to yield many other interesting results. We conclude this thesis by pointing to some other results that can potentially emerge from this connection between fairness and complexity theory. Some of these future directions are part of our work in progress, and hence we also briefly describe our current thoughts on how these other new results can be obtained.

8.1 Multi-class multicalibration

As explained in Chapter 5, in proving our PAME++ theorem we are actually using a more restrictive version of the PAME theorem shown in the work of Vadhan & Zheng [VZ13]. In particular, our PAME++ theorem should be generalized in two different ways. First, it is not necessary that we set $X = \mathcal{U}_n$ and $B = g(X)$, and our same proofs should go through for an arbitrary joint distribution (X, B). In that case, the assumption that g is δ-weakly hard would become the assumption that B is $(1 - 2^r)$-hard to predict.

Second, the PAME theorem of Vadhan & Zheng works for a joint distribution (X, B) on the domain $\{0,1\}^n \times \{0,1\}^\ell$ for $\ell = O(\log n)$, whereas we restricted our PAME++ results to the case where $\ell = 1$ (i.e., B is boolean). The reason why we chose to do so is because $\ell = 1$ allows us to use the original notion of multicalibration, whereas going into the case where $\ell > 1$ would require using a multi-class version of the notion of multicalibration. Namely, all throughout our thesis, we have assumed that a multicalibrated predictor is always real-valued. However, it makes sense to consider the notion of multicalibration in the multi-class setting; namely, when the outputs of the predictor h correspond to ℓ-bit vectors, that is, $h\colon \mathcal{X} \to \{0,1\}^\ell$. Such a notion of multicalibration has not yet received much attention in the literature, with the exception of the works on omnipredictors [GKR+21] and on low-degree multicalibration [GKSZ22]. In the former, multi-class multicalibration is defined using the notion of covariance, whereas in the latter, multi-class multicalibration is defined using inner products. Both ideas make sense because in the multi-class setting we need to somehow transform a vector into a real number, given that h outputs an ℓ-bit vector yet the ϵ parameter in multicalibration is a real number.

We have a different notion of multi-class multicalibration in mind, which is the following one:

Definition 8.1 (Multi-class multicalibration)**.** Let $\mathcal{F}\colon \mathcal{X} \times \{0,1\}^\ell \to [0,1]$, $g\colon \mathcal{X} \times \{0,1\}^\ell \to [0,1]$. A predictor $h\colon \mathcal{X} \to \{0,1\}^\ell$ is *ϵ-multicalibrated* if for all $f \in \mathcal{F}$ and for all $v \in \text{range}(h)$,

$$\left| \mathop{\mathbb{E}}_{x \sim X,\, b \sim \{0,1\}^\ell} [f(x,b) \cdot (g(x,b) - h(x,b)) \cdot \mathbb{1}[h(x,b) = v]] \right| \leq \epsilon.$$

(From this definition, one would then consider the corresponding approximate MC relaxation, as we did in the case where h is real-valued in Chapter 2.) We suspect that this definition of multi-class multicalibration is equivalent to the low-degree multicalibration notion of [GKSZ22]. Either way, our hope is that Definition 8.1 would allow us to generalize our PAME++ theorem to the setting where (X, B) is a joint distribution over $\{0,1\}^n \times \{0,1\}^\ell$. One remark about this approach is that it would require us to re-prove the PAME++ theorem by starting from a multicalibrated partition directly, rather than proving it using IHCL++. This is because it does not make sense

definitionally to generalize IHCL to the multi-class setting, given that in it we need to compute the probability that $g(x) = h(x)$, and g is a real-valued function.

Lastly, another generalization that one could do to our PAME++ results is to consider the uniform setting. That is, in the original work of Vadhan & Zheng [VZ13], they consider both the non-uniform and uniform versions of the notion of PAME, and consequently their PAME theorem is stated both in the non-uniform and uniform settings. Instead, in Chapter 5, we only considered the non-uniform setting, and therefore we expect our results to carry on into the uniform setting as well.

8.2 Distributional zero-knowledge

We remark that this section provides a high-level overview of how the paradigm of this thesis can be applied to cryptographic settings, and therefore we do not include all of the necessary definitions in this section. For the full definitions, we defer the reader to [CLP15] and [JP14].

Another reason for our interest in the notion of multi-class multicalibration is that we would also need this generalization of the notion of multicalibration in order to obtain stronger cryptographic results that are related to the regularity lemma of Trevisan et al. by applying our main paradigm. Namely, Chung et al. [CLP15] showed that their notion of distributional zero-knowledge in cryptography is much related to the regularity lemma of Trevisan et al. [TTV09]. Formally, Chung et al. consider relaxations of the original notion of zero-knowledge in cryptography and study when these relaxations are equivalent to their weak counter-parts [CLP15]. By "weak counter-parts" they mean that the order of quantifiers in the notion of zero-knowledge is switched. That is, in the original notion of zero-knowledge, we need to find a universal simulator S that is able to fool all distinguishers D. Instead, in the notion of weak zero-knowledge, we allow the simulator S to be distinguisher-dependent. Formally:

Definition 8.2 (Weak zero-knowledge, Definition 8 in [CLP15])**.** Let (P, V) be an interactive proof system for a language L. We say that (P, V) is *weak zero-knowledge* if for every PPT verifier V^* and for every PPT distinguisher D, there exists a PPT simulator S and a negligible function $\rho(\cdot)$ such that for every $n \in \mathbb{N}$, $x \in L \cap \{0,1\}^N$, and $z \in \{0,1\}^*$, we have

$$|\Pr[D(x, z, \mathrm{Out}_{V^*}[P(x) \leftrightarrow V^*(x, z)]) = 1] - \Pr[D(x, z, S(x, z)) = 1]| \leq \rho(n).$$

(For the definitions of the terms used in this definition, see [CLP15].) Chung et al. then define the following relaxation of zero-knowledge, which they prove is equivalent to its weak counter-part:

Definition 8.3 (Distributional (T, t, ϵ)-zero-knowledge, Definition 15 in [CLP15])**.** Let (P, V) be an interactive proof system for a language L. We say that (P, V) is *distributional* (T, t, ϵ)*-zero-knowledge* if for every $n \in \mathbb{N}$, every joint distribution (X_n, Y_n, Z_n) over $(L \cap \{0,1\}^n) \times \{0,1\}^* \times \{0,1\}^*$, and every $t(n)$-size verifier, there exists a randomized $T(n)$-size simulator S such that for every randomized $t(n)$-size distinguisher D, we have

$$|\Pr[D(X_n, Z_n, \mathrm{Out}_{V^*}[P(X_n, Y_n) \leftrightarrow V^*(X_n, Z_n)]) = 1] - \Pr[D(X_n, Z_n, S(X_n, Z_n)) = 1]| \leq \epsilon(n).$$

Given the equivalence between this notion and its weak counter-part, Chung et al. then prove the following theorem:

Theorem 8.4 (A laconic prover implies distributional ZK; Theorem 21 in [CLP15])**.** *Let (P, V) be an interactive proof system for a language L, and suppose that the prover P has communication complexity $\ell(n)$; i.e., the total length of the messages sent by P is $\ell(n)$, where n is the length of the common input x. Then, for every function $t'(n) \geq \Omega(n)$ and $\epsilon'(n)$, (P, V) is distributional (T', t', ϵ')-zero-knowledge, where*

$$T'(n) = O\left(2^{\ell(n)} \cdot \frac{t'(n)^3 \ln(t'(n))}{\epsilon'(n)^4}\right).$$

In the literature, this theorem is sometimes called the *interactive regularity lemma*, given that Chung et al. show that this theorem implies a version of the regularity lemma of Trevisan et al. [TTV09], and they also show how to prove the Dense Model Theorem using Theorem 8.4.

But the direction that we are interested in for future work is the one considered in the paper by Jetchev and Pietrazk [JP14], where they show that the regularity lemma implies Theorem 8.4. In particular, Jetchev and Pietrazk prove the following variant of the regularity lemma:

Theorem 8.5 ([JP14])**.** *Consider a joint distribution (X, A) on a set $\mathcal{X} \times \{0,1\}^\ell$. For any family $\mathcal{F}$ of distinguishers $f \colon \mathcal{X} \times \{0,1\}^\ell \to \{0,1\}$, there exists a simulator $h \colon \mathcal{X} \to \{0,1\}^\ell$ such that*

1. *no function in $\mathcal{F}$ can distinguish (X, A) from $(X, h(X))$ with advantage ϵ, and*
2. *h is only $O(2^{3\ell}\epsilon^{-2})$ times less efficient than the functions in $\mathcal{F}$.*

We remark that we can translate between the [JP14] and the [TTV09] formulations of the regularity lemma (incurring a loss 2^ℓ in efficiency) as follows: Given a predictor $\tilde{h} \colon \mathcal{X} \to [0,1]$, we transform it into a predictor of the form $h \colon \mathcal{X} \to \{0,1\}^\ell$ by setting $\Pr_{[h]}[h(x) = b] = \tilde{h}(x, b)$, where b is drawn from $\{0,1\}^\ell$.

On a high level, the proof that the regularity lemma of [JP14] implies Theorem 8.5 by Chung et al. follows by instantiating the domain $\mathcal{X}$ as (X_n, Y_n, Z_n), the distinguishers $\mathcal{F}$ as circuits of size t, and the arbitrary function g as $g(X_n, Y_n, Z_n) = M$, where $M \in \{0,1\}^\ell$ denotes the messages sent by the (laconic) prover $P(X_n, Y_n)$ when interacting with the verifier $V^*(X_n, Z_n)$. Then, we apply the regularity lemma to obtain a simulator h that satisfies the required indistinguishability guarantee, and from this h we build a zero-knowledge simulator S that satisfies the notion of distributional zero-knowledge (Definition 8.3). The fundamental idea is that we can build a zero-knowledge simulator that satisfies the definition of distributional zero-knowledge from a multiaccurate predictor h.

Therefore, this type of implication falls exactly into the paradigm that we have studied in this thesis: namely, the work of [JP14] shows that we can prove the key theorem of [CLP15] on interactive proof systems and distributional zero-knowledge (Theorem 8.4) using a multiaccurate predictor. Therefore, the question becomes: If we start with a multicalibrated predictor instead, what stronger theorem do we obtain? In this case, due to the nature of Theorem 8.4, we suspect that starting with a multicalibrated predictor will yield a stronger version of the definition of distributional zero-knowledge, such that an interactive proof system with a laconic prover can achieve this stronger notion. Because in this case the predictor h takes n-bit strings as inputs (given that in the definition of distributional zero-knowledge we are considering the joint distribution (X_n, Y_n, Z_n) over $(L \cap \{0,1\}^n) \times \}0,1\}^* \times \{0,1\}^*$), this is why we will need the more general notion of multi-class multicalibration in order to achieve a stronger distributional zero-knowledge definition using the paradigm of this thesis. We believe that this would be a very interesting result to achieve, given that it would extend our paradigm to the field of cryptography.

8.3 Other future directions

Another potential future direction in the setting of cryptography is that of *leakage resilient cryptography*. Namely, in the work of Jetchev and Pietrzak, the original motivation is to show how to use their variant of the regularity lemma (namely, Theorem 8.5) to simplify the security proofs of leakage-resilient and other cryptosystems whose security proofs rely on chain rules for computational entropy. Namely, in their setting, we think of the $\{0,1\}^\ell$ term in their Theorem 8.5 as a short ℓ-bit auxiliary input to the distinguishers. As summarized in [JP14], the idea is to replace the chain rules simulation-based arguments, where the required indistinguishability is provided by the regularity lemma. Namely, Theorem 8.5 shows that the leakage $\{0,1\}^\ell$ can be "efficiently simulated". Hence, this is another example of an implication that falls into our paradigm: That is, we can use a multiaccurate predictor to prove certain chain rules for computational entropy (following the proofs and ideas given in [JP14]). Then, if we start with a multicalibrated predictor instead, what stronger implication would be obtain?

Outside of cryptography, another possible connection that falls into this paradigm is one that relates the DMT and pseudoentropy. More concretely, in the work of Reingold et al., they characterize the Dense Model Theorem in terms of pseudoentropy [RTTV08]. Given our DMT++ and PAME++ statements, we believe that it would be interesting to analyze the connection pointed out in [RTTV08]. For example, if we plug in our DMT++ into their characterization of DMT as pseudoentropy, do we recover our PAME++ theorem, or do we obtain some other variant of it?

Lastly, another interesting research direction is to study how our ++ theorems change when considering different notions of multicalibration. For example, Gopalan et al. recently proposed a new relaxation of multicalibration, called *swap multicalibration*, which is a stronger notion than the approximate multicalibration definition that we have used in this thesis, yet it is also efficiently realizable [GKR23]. We believe that using swap multicalibration instead of approximate multicalibration in our ++ theorems would allow us to obtain better parameters.

Glossary

Notations

$\mathcal{F}$	set of distinguishers
f	distinguisher
g	arbitrarily complex function
h	simulator/predictor
$\mathcal{X}$	domain (a finite set)
$\mathcal{D}$	probability distribution
$\mathcal{D}\|_P$	conditional distribution for P
$\mathbb{E}$	expectation
$x \sim \mathcal{D}$	x is sampled from $\mathcal{D}$
Pr	probability
X_v	level set of h at v
Λ	λ-discretization
$\mathcal{P}$	partition of $\mathcal{X}$
η_p	size of P
v_p	expectation of g on X_v
k_p	balance of g on P
γ	lower bound on η_p
τ	lower bound on k_p
$e_G(S,T)$	number of edges between S and T
$d_G(S,T)$	density for S and T
$\mathcal{U}_\mathcal{X}$	uniform distribution over $\mathcal{X}$
μ	measure on $\mathcal{X}$
$\mathcal{D}_\mu$	distribution for a measure μ
$\mathbb{1}$	indicator random variable
$\mathbb{1}_G$	indicator for (γ,τ)-good partition sets
$\mathrm{Bern}(v)$	Bernoulli random variable of parameter v
H	Shannon entropy
H_∞	min-entropy

Abbreviations

MA	multiaccurate
MC	multicalibrated
OI	outcome indistinguishability
IHCL	Impagliazzo's Hardcore Lemma
IHCL++	generalized Impagliazzo's Hardcore Lemma
DMT	Dense Model Theorem
DMT++	generalized Dense Model Theorem
PAME	pseudo-average min-entropy (theorem)
PAME++	generalized PAME (theorem)

Bibliography

[Bar20] Boaz Barak. *An Intensive Introduction to Cryptography*. Creative Commons, 2020.

[Bar22] Boaz Barak. *Introduction to Theoretical Computer Science*. Creative Commons, 2022.

[BG18] Joy Buolamwini and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification". In: *Conference on fairness, accountability and transparency*. PMLR. 2018, pp. 77–91.

[BHN17] Solon Barocas, Moritz Hardt, and Arvind Narayanan. "Fairness in machine learning". In: *Nips tutorial* 1 (2017), p. 2.

[BSW03] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. "Computational analogues of entropy". In: *Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques: 6th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2003 and 7th International Workshop on Randomization and Approximation Techniques in Computer Science, RANDOM 2003, Princeton, NJ, USA, August 24-26, 2003. Proceedings*. Springer. 2003, pp. 200–215.

[BYR+21] Noam Barda, Gal Yona, Guy N Rothblum, Philip Greenland, Morton Leibowitz, Ran Balicer, Eitan Bachmat, and Noa Dagan. "Addressing bias in prediction models by improving subpopulation calibration". In: *Journal of the American Medical Informatics Association* 28.3 (2021), pp. 549–558.

[CCL18] Yi-Hsiu Chen, Kai-Min Chung, and Jyun-Jie Liao. "On the complexity of simulating auxiliary input". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2018, pp. 371–390.

[Cho17] Alexandra Chouldechova. "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments". In: *Big data* 5.2 (2017), pp. 153–163.

[CLP15] Kai-Min Chung, Edward Lui, and Rafael Pass. "From weak to strong zero-knowledge and applications". In: *Theory of Cryptography Conference*. Springer. 2015, pp. 66–92.

[Cub17] Maria Ramon Cubells Bartolomé. "La irreductible percepció sensible i l'apreciació de la bellesa artística en Leibniz". In: *Enrahonar* 59 (2017), pp. 111–127.

[Daw85] A Philip Dawid. "Calibration-based empirical probability". In: *The Annals of Statistics* 13.4 (1985), pp. 1251–1274.

[DHP+12] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. "Fairness through awareness". In: *Proceedings of the 3rd innovations in theoretical computer science conference*. 2012, pp. 214–226.

[DKR+21] Cynthia Dwork, Michael P Kim, Omer Reingold, Guy N Rothblum, and Gal Yona. "Outcome indistinguishability". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 1095–1108.

[DLLT23] Cynthia Dwork, Daniel Lee, Huijia Lin, and Pranay Tankala. "New Insights into Multi-Calibration". In: *arXiv preprint arXiv:2301.08837* (2023).

[DRS04] Yevgeniy Dodis, Leonid Reyzin, and Adam Smith. "Fuzzy extractors: How to generate strong keys from biometrics and other noisy data". In: *Advances in Cryptology-EUROCRYPT 2004: International Conference on the Theory and Applications of Cryptographic Techniques, Interlaken, Switzerland, May 2-6, 2004. Proceedings 23*. Springer. 2004, pp. 523–540.

[Fel09] Vitaly Feldman. "Distribution-specific agnostic boosting". In: *arXiv preprint arXiv:0909.2927* (2009).

[FK99] Alan Frieze and Ravi Kannan. "Quick approximation to matrices and applications". In: *Combinatorica* 19.2 (1999), pp. 175–220.

[FS99] Yoav Freund and Robert E Schapire. "Adaptive game playing using multiplicative weights". In: *Games and Economic Behavior* 29.1-2 (1999), pp. 79–103.

[GHK+22] Parikshit Gopalan, Lunjia Hu, Michael P Kim, Omer Reingold, and Udi Wieder. "Loss minimization through the lens of outcome indistinguishability". In: *arXiv preprint arXiv:2210.08649* (2022).

[GHK+23] Ira Globus-Harris, Declan Harrison, Michael Kearns, Aaron Roth, and Jessica Sorrell. "Multicalibration as Boosting for Regression". In: *arXiv preprint arXiv:2301.13767* (2023).

[GIK12] Sitanshu Gakkhar, Russell Impagliazzo, and Valentine Kabanets. "Hardcore measures, dense models and low complexity approximations". PhD thesis. Simon Fraser University, 2012.

[GKR+21] Parikshit Gopalan, Adam Tauman Kalai, Omer Reingold, Vatsal Sharan, and Udi Wieder. "Omnipredictors". In: *arXiv preprint arXiv:2109.05389* (2021).

[GKR23] Parikshit Gopalan, Michael P Kim, and Omer Reingold. "Characterizing notions of omniprediction via multicalibration". In: *arXiv preprint arXiv:2302.06726* (2023).

[GKSZ22] Parikshit Gopalan, Michael P Kim, Mihir A Singhal, and Shengjia Zhao. "Low-degree multicalibration". In: *Conference on Learning Theory*. PMLR. 2022, pp. 3193–3234.

[GM82] Shafi Goldwasser and Silvio Micali. "How To Play Mental Poker Keeping Secret". In: (1982).

[GNW11] Oded Goldreich, Noam Nisan, and Avi Wigderson. "On Yao's XOR-Lemma." In: *Studies in Complexity and Cryptography* 6650 (2011), pp. 273–301.

[Gow10] W Timothy Gowers. "Decompositions, approximate structure, transference, and the Hahn–Banach theorem". In: *Bulletin of the London Mathematical Society* 42.4 (2010), pp. 573–606.

[Gow97] William T Gowers. "Lower bounds of tower type for Szemerédi's uniformity lemma". In: *Geometric & Functional Analysis GAFA* 7.2 (1997), pp. 322–337.

[GRSW22] Parikshit Gopalan, Omer Reingold, Vatsal Sharan, and Udi Wieder. "Multicalibrated partitions for importance weights". In: *International Conference on Algorithmic Learning Theory*. PMLR. 2022, pp. 408–435.

[GT08] Ben Green and Terence Tao. "The primes contain arbitrarily long arithmetic progressions". In: *Annals of mathematics* (2008), pp. 481–547.

[GW11] Craig Gentry and Daniel Wichs. "Separating succinct non-interactive arguments from all falsifiable assumptions". In: *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 2011, pp. 99–108.

[HF22] Emmie Hine and Luciano Floridi. "The Blueprint for an AI Bill of Rights: in search of enaction, at risk of inaction". In: *Available at SSRN* (2022).

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. "A pseudorandom generator from any one-way function". In: *SIAM Journal on Computing* 28.4 (1999), pp. 1364–1396.

[HKRR18] Ursula Hébert-Johnson, Michael Kim, Omer Reingold, and Guy Rothblum. "Multicalibration: Calibration for the (computationally-identifiable) masses". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 1939–1948.

[HLR07] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. "Conditional computational entropy, or toward separating pseudoentropy from compressibility". In: *Advances in Cryptology-EUROCRYPT 2007: 26th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Barcelona, Spain, May 20-24, 2007. Proceedings 26*. Springer. 2007, pp. 169–186.

[Hol05] Thomas Holenstein. "Key agreement from weak bit agreement". In: *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*. 2005, pp. 664–673.

[Ilv19] Christina Ilvento. "Metric learning for individual fairness". In: *arXiv preprint arXiv:1906.00250* (2019).

[Imp08] Russell Impagliazzo. "When Do Sparse Sets Have Dense Models". In: *Pseudorandomness in Mathematics and Computer Science Miniworkshop* (2008).

[Imp09] Russell Impagliazzo. "Algorithmic Dense Model Theorems and Weak Regularity". In: (2009).

[Imp95] Russell Impagliazzo. "Hard-core distributions for somewhat hard problems". In: *Proceedings of IEEE 36th Annual Foundations of Computer Science*. IEEE. 1995, pp. 538–545.

[IMR14] Russell Impagliazzo, Cristopher Moore, and Alexander Russell. "An entropic proof of Chang's inequality". In: *SIAM Journal on Discrete Mathematics* 28.1 (2014), pp. 173–176.

[JP14] Dimitar Jetchev and Krzysztof Pietrzak. "How to fake auxiliary input". In: *Theory of Cryptography Conference*. Springer. 2014, pp. 566–590.

[Kal04] Adam Kalai. "Learning monotonic linear functions". In: *LECTURE NOTES IN COMPUTER SCIENCE.* (2004), pp. 487–501.

[KGZ19] Michael P Kim, Amirata Ghorbani, and James Zou. "Multiaccuracy: Black-box post-processing for fairness in classification". In: *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. 2019, pp. 247–254.

[Kim20] Michael Pum-Shin Kim. *A complexity-theoretic perspective on fairness*. Stanford University, 2020.

[KK09] Varun Kanade and Adam Kalai. "Potential-based agnostic boosting". In: *Advances in neural information processing systems* 22 (2009).

[KKG+22] Michael P Kim, Christoph Kern, Shafi Goldwasser, Frauke Kreuter, and Omer Reingold. "Universal adaptability: Target-independent inference that competes with propensity scoring". In: *Proceedings of the National Academy of Sciences* 119.4 (2022), e2108097119.

[KM96] Michael Kearns and Yishay Mansour. "On the boosting ability of top-down decision tree learning algorithms". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 1996, pp. 459–468.

[KMV08] Adam Tauman Kalai, Yishay Mansour, and Elad Verbin. "On agnostic boosting and parity learning". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. 2008, pp. 629–638.

[KNRW18] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. "Preventing fairness gerrymandering: Auditing and learning for subgroup fairness". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 2564–2572.

[Kra18] Alexander Krauss. "Why all randomised controlled trials produce biased results". In: *Annals of medicine* 50.4 (2018), pp. 312–322.

[KS03] Adam R Klivans and Rocco A Servedio. "Boosting and hard-core set construction". In: *Machine Learning* 51.3 (2003), pp. 217–238.

[Lee17] Holden Lee. "Learning Models of Mathematical Objects". In: (2017).

[LMKA16] Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin. "How we analyzed the COMPAS recidivism algorithm". In: *ProPublica (5 2016)* 9.1 (2016), pp. 3–3.

[LWM22] Johann Laux, Sandra Wachter, and Brent Mittelstadt. "Trustworthy Artificial Intelligence and the European Union AI Act: On the Conflation of Trustworthiness and the Acceptability of Risk". In: *Available at SSRN 4230294* (2022).

[MM02] Yishay Mansour and David McAllester. "Boosting using branching programs". In: *Journal of Computer and System Sciences* 64.1 (2002), pp. 103–112.

[NR23] Georgy Noarov and Aaron Roth. "The Scope of Multicalibration: Characterizing Multicalibration via Property Elicitation". In: *arXiv preprint arXiv:2302.08507* (2023).

[ONe17] Cathy O'Neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, 2017.

[RTTV08] 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*. IEEE. 2008, pp. 76–85.

[Skó15] Maciej Skórski. "Efficiently simulating high min-entropy sources in the presence of side information". In: *International Conference on Cryptology in India*. Springer. 2015, pp. 312–325.

[Sko16] Maciej Skorski. "Simulating auxiliary inputs, revisited". In: *Theory of Cryptography Conference*. Springer. 2016, pp. 159–179.

[Skó16] Maciej Skórski. "A subgradient algorithm for computational distances and applications to cryptography". In: *Cryptology ePrint Archive* (2016).

[Skó17] Maciej Skórski. "A cryptographic view of regularity lemmas: Simpler unified proofs and refined bounds". In: *International Conference on Theory and Applications of Models of Computation*. Springer. 2017, pp. 586–599.

[Tao05] Terence Tao. "The dichotomy between structure and randomness, arithmetic progressions, and the primes". In: *arXiv preprint math/0512114* (2005).

[Tao07] Terence Tao. "Structure and randomness in combinatorics". In: *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE. 2007, pp. 3–15.

[TTV09] Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. "Regularity, boosting, and efficiently simulating every high-entropy distribution". In: *2009 24th Annual IEEE Conference on Computational Complexity*. IEEE. 2009, pp. 126–136.

[TZ08] Terence Tao and Tamar Ziegler. "The primes contain arbitrarily long polynomial progressions". In: *Acta Mathematica* 201.2 (2008), pp. 213–305.

[Vin18] James Vincent. "Amazon reportedly scraps internal AI recruiting tool that was biased against women". In: *The Verge* 10 (2018).

[VZ12] 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*. 2012, pp. 817–836.

[VZ13] Salil Vadhan and Colin Jia Zheng. "A uniform min-max theorem with applications in cryptography". In: *Annual Cryptology Conference*. Springer. 2013, pp. 93–110.

[WAB+19] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. "Disability, bias, and AI". In: *AI Now Institute* (2019).

[Yao82] Andrew C Yao. "Theory and application of trapdoor functions". In: *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)*. IEEE. 1982, pp. 80–91.

[Zhe14] Jia Zheng. *A uniform min-max theorem and characterizations of computational randomness*. Harvard University, 2014.

This thesis was typeset using LaTeX, originally developed by Leslie Lamport and based on Donald Knuth's TeX. The above illustration, *Science Experiment 02*, was created by Ben Schlitter and released under CC BY-NC-ND 3.0. A template that can be used to format a PhD dissertation with this look *&* feel has been released under the permissive AGPL license, and can be found online at github.com/suchow/Dissertate or from its lead author, Jordan Suchow, at suchow@post.harvard.edu.