Replicability in High Dimensional Statistics

Max Hopkins
Computer Science and Engineering
University of California San Diego
La Jolla, USA
nmhopkin@ucsd.edu

Russell Impagliazzo
Computer Science and Engineering
University of California San Diego
La Jolla, USA
rimpagliazzo@ucsd.edu

Daniel Kane
Computer Science and Engineering
University of California San Diego
La Jolla, USA
dakane@ucsd.edu

Sihan Liu
Computer Science and Engineering
University of California San Diego
La Jolla, USA
sil046@ucsd.edu

Christopher Ye
Computer Science and Engineering
University of California San Diego
La Jolla, USA
czye@ucsd.edu

2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS) | 979-8-3315-1674-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/FOCS61266.2024.00103

***Abstract*—The replicability crisis is a major issue across nearly all areas of empirical science, calling for the formal study of replicability in statistics. Motivated in this context, [Impagliazzo, Lei, Pitassi, and Sorrell STOC 2022] introduced the notion of replicable learning algorithms, and gave basic procedures for 1-dimensional tasks including statistical queries. In this work, we study the computational and statistical cost of replicability for several fundamental *high dimensional* statistical tasks, including multi-hypothesis testing and mean estimation.**

Our main contribution establishes a computational and statistical equivalence between optimal replicable algorithms and high dimensional isoperimetric tilings. As a consequence, we obtain matching sample complexity upper and lower bounds for replicable mean estimation of distributions with bounded covariance, resolving an open problem of [Bun, Gaboardi, Hopkins, Impagliazzo, Lei, Pitassi, Sivakumar, and Sorrell, STOC 2023] and for the N-Coin Problem, resolving a problem of [Karbasi, Velegkas, Yang, and Zhou, NeurIPS 2023] up to log factors.

While our equivalence is computational, allowing us to shave log factors in sample complexity from the best known efficient algorithms, efficient isoperimetric tilings are not known. To circumvent this, we introduce several relaxed paradigms that do allow for sample and computationally efficient algorithms, including allowing pre-processing, adaptivity, and approximate replicability. In these cases we give efficient algorithms matching or beating the best known sample complexity for mean estimation and the coin problem, including a generic procedure that reduces the standard quadratic overhead of replicability to linear in expectation.

***Index Terms*—replicability, high dimensional statistics, foams, isoperimetry, learning theory**

I. Introduction

The *replicability crisis* permeates almost all areas of science. Recent years have seen the repeated failure of influential work in oncology [1], clinical research [2], and other high impact areas to replicate under scrutiny. Indeed the problem is so pervasive that in a survey of 1500 scientists, 70% reported they had tried and failed to replicate another researcher's findings [3]. While many factors underlie the failure of replicability in science, a key component is the instability of underlying *statistical methods*. Even techniques as basic as hypothesis testing suffer from these issues [4], and combined with the explosion in *number* of performed tests each year, it seems inevitable published false positives will skyrocket unless new methods are developed.

Motivated in this context, we study the cost of replicability in statistics in the recent algorithmic framework of Impagliazzo, Lei, Pitassi, and Sorrell [5]. An algorithm $\mathcal{A}$ drawing samples from an (unknown) population $\mathcal{D}$ is called ρ-replicable if, run twice on *independent* samples and *the same* randomness, $\mathcal{A}$ produces exactly the same answer with probability $1-\rho$. We focus on characterizing the computational and statistical complexity of replicability for two core interrelated problems: *multi-hypothesis testing* and *high dimensional mean estimation*.

As a warm-up, consider the setting of a single hypothesis test. A typical procedure sets up a test statistic Z to distinguish between a null $h_\emptyset$ and alternative hypothesis h_1 such that under $h_\emptyset$, Z is uniform on $[0,1]$, while under h_1 there exists $q_0 > p_0$ such that $\Pr[Z \le p_0] \ge q_0$. Formalized in this way, hypothesis testing is equivalent to one of the earliest problems in replicability and distribution testing, the *coin problem* (testing the bias of a weighted coin). Despite its central position, the complexity of the replicable coin problem is not fully understood. Worse, current methods have *quadratic* overhead in ρ which may be infeasible in practice. Our first contribution is a tight characterization of the coin problem, reducing this cost to just linear in expectation.

The coin problem is a fundamental example of *1-dimensional problem* in statistics but, in practice, most problems are really *high dimensional*. An epidemiologist, may, for instance, want to test the prevalence of a suite of N diseases in some population. Or, even in a single hypothesis test, the test statistic itself may involve computing the mean of some N-dimensional data; if such pre-processing steps are non-

M. Hopkins, D. Kane, and S. Liu are supported by NSF Award CCF-1553288 (CAREER) and a Sloan Research Fellowship. R. Impagliazzo and C. Ye are supported by NSF Award AF: Medium 2212136. C. Ye is additionally supported by NSF grants 1652303, 1909046, 2112533, and HDR TRIPODS Phase II grant 2217058.

2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)

2575-8454/24/$31.00 ©2024 IEEE
DOI 10.1109/FOCS61266.2024.00103

replicable, the final test may be as well. This brings us to the main question addressed in this paper: *how does the cost of replicability scale with dimension N?*

High dimensional replicability in this sense was first considered in [6] and [7]. In [7], the authors study the N-*Coin Problem*, akin to the 'multi-hypothesis' setup above. They argue that while independently estimating each coin replicably takes N^3 flips, by *correlating* choices one can improve this cost to N^2, albeit in exponential time. Likewise, [6] show a correlated strategy for replicably estimating an N-dimensional Gaussian in N^2 samples. At outset, it was unclear whether the proposed strategies were optimal: while [7] conjectured no better algorithm could exist, [6] asked if the problem could be solved in N samples. Is there a principled approach to understanding the cost of such problems?

We resolve this question by proving a tight connection between high dimensional replicability and a well-studied problem in high dimensional geometry: *low surface area tilings of* $\mathbb{R}^N$. Low surface area tilings, closely related to optimal packings, are a classical problem dating back to Pappus of Alexandria in the 4th century,[1] with asymptotically optimal constructions known since the 1950s [9], [10]. In computer science, such tilings have seen more recent study due to their close connections with lattice cryptography (see e.g. [11], [12]) and hardness of approximation [13], [14].

We prove a computational and statistical equivalence between (efficient) replicable algorithms and (efficient) tilings. Given a replicable algorithm with low sample complexity, we give an oracle-efficient construction of an (approximate) tiling with low surface area. Conversely given an (approximate) tiling with low surface area, we give an oracle-efficient replicable algorithm with low sample complexity. Applying the classical isoperimetric theorem, we immediately get near-tight lower bounds for Gaussian mean estimation and the N-Coin Problem matching the algorithms of [6], [7] up to log factors, resolving their corresponding open questions.[2]

On the algorithmic side, while isoperimetric tilings exist, all known constructions take exponential time. Thus achieving true sample optimality via this approach, similar to [6], [7], currently requires exponential time. On the other hand, there are efficient tilings that (slightly) beat naive 'independent estimation' [11]. Combined with our equivalence theorem, this gives the best known polynomial time algorithms for N-dimensional mean estimation and the coin problem. Further, even if no efficient isoperimetric tilings exist, we argue it is nevertheless possible to *pre-process* an inefficient tiling in such a way that sample-optimal replicability can be achieved in polynomial time with query access to the pre-processing output. We leave the construction (or hardness of) truly efficient isoperimetric tilings as the main question (re)raised by this work.

[1]Pappus claimed a solution for the 2-dimensional case, later proved by Hales [8].

[2]Formally, we resolve the sample complexity of the *non-adaptive* N-Coin problem up to log factors. The authors of [7] do not consider the adaptive sample model. We discuss this subtlety later on.

Finally, in light of the lack of efficient isoperimetric tilings, we introduce two relaxed paradigms for replicability and multi-hypothesis testing that do allow for sample and computationally efficient algorithms. First, we consider *adaptive* algorithms which may choose which of N coins they flip during execution based on prior observations. We exhibit a polynomial time algorithm in this model matching the best-known sample complexity of prior (inefficient) non-adaptive methods. Second, we look at relaxations that only require *approximate* replicability. In particular, we show if one only requires the outputs over two runs to agree on most coins, it is possible to build efficient algorithms *beating* the sample complexity implied by isoperimetric tilings.

A. Our Contributions

Before stating our results, we briefly recall the formal notion of a replicable algorithm.

Definition 1.1 (Impagliazzo, Lei, Pitassi, and Sorrell [5]): An algorithm $\mathcal{A}$ is ρ-replicable if for all distributions $\mathcal{D}$ and i.i.d. samples $S, S' \sim \mathcal{D}$

$$\Pr_{r,S,S'}\left(\mathcal{A}(S;r) = \mathcal{A}(S';r)\right) \geq 1-\rho,$$

where r denotes the internal randomness of the algorithm $\mathcal{A}$. Replicable algorithms are inherently randomized, and typically have a corresponding 'failure probability' δ. For simplicity, in this overview we will ignore sample dependence on δ which always scales *logarithmically* in $\frac{1}{\delta}$. Formally, the below results can be thought of as in the regime where $\delta = \Theta(\rho)$. Formal dependencies on all parameters are given in the main body.

1) On Replicability in 1*-Dimension:* While our eventual goal is to understand the price of replicability in high dimensions, it is of course natural to first ask for a tight understanding in 1-dimension. With this in mind, we first consider the fundamental problems of *single hypothesis testing* and *bias estimation*.

Suppose we have some hypothesis h_0 and an experiment designed to test this hypothesis is repeated m times, thus creating a sequence of m p-values. If h_0 is true, then the p-values should be uniformly distributed. On the other hand, if h_0 is false, we should gather small p-values with higher probability than normal. Quantitatively, there are constants p_0, q_0 such that p-values smaller than p_0 are observed with probability $q_0 > p_0$ (in statistics, q_0 is called the *power* of the experiment). Given a sequence of p-values, we want to design an algorithm that *replicably* determines whether to reject the null hypothesis h_0. We formalize this in the following definition.

Definition 1.2 (Hypothesis Testing): Let $0 \leq p_0 < q_0 \leq 1$, $\delta < \frac{1}{2}$. A (randomized) algorithm $\mathcal{A}$ is a (p_0, q_0)-hypothesis tester if given sample access S to some unknown $\mathcal{D}$ on $[0,1]$:

1) Given $\mathcal{D} = \text{Unif}([0,1])$, then $\Pr(\mathcal{A}(S) = \textsc{Reject}) < \delta$.
2) Given $\Pr_{x\sim\mathcal{D}}(x < p_0) \geq q_0$, then $\Pr(\mathcal{A}(S) = \textsc{FailToReject}) < \delta$.

Single hypothesis testing is computationally and statistically equivalent to a well-studied problem in distribution testing,

the *coin problem* [15]. Given a coin with a hidden bias p, the (p_0, q_0)-coin problem asks the learner to determine whether the bias p is at most p_0, or at least q_0. The coin problem was one of the first questions studied in algorithmic replicability and plays a critical role as a subroutine in later works. Nevertheless, there is a still gap in the best known bounds:

Theorem 1.3 (Impagliazzo, Lei, Pitassi, and Sorrell [5], Karbasi, Velegkas, Yang, and Zhou [7]): Let $p_0, q_0 \in (0, 1/2)$ and $\rho \in (0, 1)$, there is a computationally efficient ρ-replicable algorithm for the (p_0, q_0)-coin problem using

$$\tilde{O}\left(\frac{1}{(q_0 - p_0)^2\rho^2}\right)$$

samples. Conversely, any algorithm for the (p_0, q_0)-coin problem uses at least

$$\Omega\left(\frac{p_0}{(q_0 - p_0)^2\rho^2}\right)$$

samples in the worst-case.

We tighten Theorem 1.3 in two key aspects. First, we resolve the gap in sample dependence on p_0 and q_0 in the numerator. Second, we address a more subtle issue regarding Theorem 1.3's dependence on ρ. In particular, we argue that while quadratic dependence on ρ is indeed necessary in the worst-case, *in expectation* the dependence can actually be reduced to *linear*.

Theorem 1.4: Let $p_0, q_0 \in (0, 1/2)$ and $\rho \in (0, 1)$. The ρ-replicable (p_0, q_0)-coin problem coin problem requires

$$\tilde{\Theta}\left(\frac{q_0}{(q_0 - p_0)^2\rho}\right)$$

samples in expectation. Moreover, the same bound holds in the worst-case with quadratic dependence on ρ and the upper bound is computationally efficient.

A few remarks are in order. First, we note that the linear overhead of Theorem 1.4 is not specific to the coin problem. In fact, we give a generic amplification lemma showing *any* replicable procedure can be performed with linear overhead (in ρ) in expectation (see Section 6 of full paper for details). Second, we remark that as an immediate consequence of Theorem 1.4 we obtain a generic procedure to efficiently transform any *non*-replicable distribution testing algorithm into a replicable one with linear expected overhead. In particular, let $\mathcal{H}_0, \mathcal{H}_1$ be two families of distributions and suppose some distribution testing algorithm $\mathcal{A}$ accepts samples from distributions $\mathcal{D} \in \mathcal{H}_0$ with probability at most $\frac{1}{3}$ and rejects samples from distributions $\mathcal{D} \in \mathcal{H}_1$ with probability at most $\frac{2}{3}$. We may view the output of $\mathcal{A}$ as a biased coin and apply Theorem 1.4 to replicably determine membership in $\mathcal{H}_0$ or $\mathcal{H}_1$ with high probability. This gives replicable algorithms for a wide range of distribution testing problems including uniformity, closeness, independence, log-concavity, and monotonicity.

For simplicity of presentation, in the rest of the introduction we state only worst-case sample complexity with quadratic dependence on ρ. Up to polylog factors, all our bounds can equivalently be stated in terms of expected complexity with linear dependence.

2) Replicability and Isoperimetry in High Dimensional Statistics: In many applications, scientists may wish to perform multiple experiments simultaneously; an epidemiologist, for instance, may want to determine the prevalence of several diseases or conditions in a population at once. Consider a setting in which a scientist runs N simultaneous hypothesis tests. In the context of replicability, we'd like to ensure that all N findings are simultaneously replicable—how does the cost of this guarantee scale with N and ρ?

Like single hypothesis testing, such a *multi*-hypothesis test is equivalent to the problem of testing biases of multiple coins (typically called the N-Coin Problem). In this section, we study the more general problem of *high dimensional mean estimation*. In particular, given sample access to a distribution $\mathcal{D}$ over $\mathbb{R}^N$, how many samples are required to ρ-replicably output an estimate $\hat{\mu}$ s.t.

$$\Pr_{S\sim D}[||\hat{\mu} - \mu_{\mathcal{D}}||_p \geq \varepsilon] \leq \delta\,?$$

We say such an algorithm (ε, ℓ_p)-learns the mean μ_D and refer to the problem of giving such an estimator as the (ε, ℓ_p)-*mean estimation problem*. We will always assume the distribution $\mathcal{D}$ has bounded covariance. Up to log factors, the N-Coin problem is the special case where $\mathcal{D}$ is the product of N independent Bernoullis and $p = \infty$ (see full paper for details).

Our core contribution is that replicable mean estimation (and therefore multihypothesis testing) is *computationally and statistically equivalent* to the construction of (approximate) low-surface area tilings of space. To state this more formally, first consider the notion of an approximate tiling:

Definition 1.5 (Isoperimetric Approximate Tilings): A (γ, A)-isoperimetric approximate tiling (IAT) of $\mathbb{R}^N$ is a collection of sets $\mathcal{P} = \{P\}$ such that for any cube $\mathcal{C} \subset \mathbb{R}^N$

1) (γ-Approximate Volume): $\text{vol}_N(\mathcal{P} \cap \mathcal{C}) \geq (1-\gamma)\text{vol}_N(\mathcal{C})$.
2) (A-Approximate Isoperimetry): $\text{vol}_{N-1}(\partial\mathcal{P} \cap \mathcal{C}) \leq A\text{vol}_N(\mathcal{C})$.
3) (Bounded Diameter): Each $P \in \mathcal{P}$ has diameter at most 1.

We call $\mathcal{P}$ **efficient** if there is an efficient membership oracle $\mathcal{O} : \mathbb{R}^N \to \mathcal{P}$ such that for any $P \in \mathcal{P}$ and $w \in P$, $\mathcal{O}(w) = P$ with high probability.

In other words, a good approximate tiling covers 'most' of $\mathbb{R}^N$ with diameter 1 bubbles with low surface-area to volume ratio. We prove the sample complexity of replicable mean estimation tightly corresponds to the surface area of an associated tiling, and moreover that there are oracle-efficient reductions between the two. We state the theorem below only for the case of ℓ_2-estimation, but will discuss its implications and variants for any $p \in [2, \infty]$ shortly.

Theorem 1.6 (Replicability $\iff$ Isoperimetry):

1) (Replicability $\to$ Isoperimetry): Let $\mathcal{A}$ be a ρ-replicable algorithm on m samples that (ε, ℓ_2)-learns the mean of N independent Bernoulli variables. Given oracle access to

$\mathcal{A}$, there is an efficient algorithm generating an efficient N-dimensional $(\rho, O(\varepsilon\rho\sqrt{m}))$-IAT.

2) (Isoperimetry $\rightarrow$ Replicability): Let $\mathcal{P}$ be an N-dimensional (ρ, A)-IAT. Given access to $\mathcal{P}$'s membership oracle and sample access to a bounded covariance ditribution $\mathcal{D}$ over $\mathbb{R}^N$, there is an efficient ρ-replicable algorithm that (ε, ℓ_2)-learns $\mu_{\mathcal{D}}$ in $O(\frac{A^2}{\varepsilon^2\rho^2})$ samples.[3]

A few remarks are in order. First, notice the surface area and sample complexity in Theorem 1.6 'match' up to constant factors. That is starting with an m-sample algorithm we get an IAT with surface area $O(\varepsilon\rho\sqrt{m})$. Starting with a surface area $O(\varepsilon\rho\sqrt{m})$-IAT, we get an algorithm on $O(\frac{(\varepsilon\rho\sqrt{m})^2}{\varepsilon^2\rho^2}) = O(m)$ samples. Second, we note the forward direction above really only relies on the family of input distributions satisfying certain mutual information bounds (see full paper for details), and therefore also holds e.g. for standard Gaussians.

By the isoperimetric inequality, the best possible surface area for an isoperimetric approximate tiling is $A = \Omega(N)$, while simply tiling space by cubes achieves $A = O(N^{3/2})$.[4] Moreover, constructions of isoperimetric tilings, that is $(0, O(N))$-IATs, have existed since the 50's [9]. Combined with Theorem 1.6, these facts lead to a tight statistical characterization of replicable mean estimation:

Corollary 1.7 (Replicable ℓ_2 Mean Estimation): Let $\varepsilon, \rho \in (0,1)$. The ρ-replicable (ε, ℓ_2)-mean-estimation problem requires

$$\Theta\left(\frac{N^2}{\varepsilon^2\rho^2}\right)$$

samples. Moreover, the lower bound holds even under Bernoulli or Gaussian distributions.

Corollary 1.7 resolves (in the negative) [6, Open Question 4] regarding whether estimation can be performed in $O(N)$ samples, as well as the ℓ_2-variant of [7]'s question regarding the complexity of the N-Coin Problem.

a) Computational Efficiency: Theorem 1.6 and Corollary 1.7 leave two important questions: what can we say about computational efficiency, and to what extent does the above hold for norms beyond ℓ_2? Toward the former, unfortunately all known isoperimetric tilings have membership oracles that run in (at best) *exponential* time, so the above algorithms are not efficient. The best known tiling with an efficient membership oracle, a lattice-based construction of Micciancio [11], only manages to shave a log factor. Nevertheless, this gives the first efficient algorithm for replicable mean estimation with (slightly) sub-cubic sample complexity.

Corollary 1.8 (Efficient Mean Estimation in Sub-Cubic Samples): Let $\varepsilon, \rho \in (0,1)$. There is an efficient ρ-replicable algorithm for (ε, ℓ_2)-mean-estimation using

$$O\left(\frac{N^3}{\rho^2\varepsilon^2} \cdot \frac{\log\log(N)}{\log(N)}\right)$$

[3]This statement assumes $\delta \geq 2^{-N}$ for simplicity. The true bound is $O\left(\frac{A^2}{\varepsilon^2\rho^2} + \frac{A^2\log\frac{1}{\delta}}{N\varepsilon^2\rho^2}\right)$.

[4]This comes from the diameter restriction. To have diameter 1, the cubes must be of side-length $\frac{1}{\sqrt{N}}$.

samples.

By the reverse direction of our reduction, any algorithm beating the above must imply improved efficient IATs. In the lattice setting, this problem has remained open since it was proposed in Micciancio's work [11]. We leave the construction of tilings satisfying our relaxed approximate notion as the main open question from this work.

b) Replicability Beyond the ℓ_2-Norm: Finally, recall in the context of hypothesis testing we are really interested in learning biases in ℓ_∞ rather than ℓ_2-norm. A version of the equivalence theorem indeed holds for general ℓ_p-norms as a consequence of the forward direction of the ℓ_2 equivalence (Theorem 1.6), an ℓ_∞ learner based on IATs (Theorem 1.10), and Hölder's inequality.

Corollary 1.9 (ℓ_p-norm Replicability $\iff$ Tilings): Fix $p \in [2, \infty], \rho \in (0,1), \varepsilon \in (0, 0.1)$. Then:

1) (Replicability $\rightarrow$ Isoperimetry): Let $\mathcal{A}$ be a $(\rho/24)$-replicable algorithm on m samples that $(\frac{\varepsilon}{N^{\frac{1}{2}-\frac{1}{p}}}, \ell_p)$-learns biases of N Bernoulli variables. Given oracle access to $\mathcal{A}$, there is an efficient algorithm generating an efficient N-dimensional $(\rho, O(\varepsilon\rho\sqrt{m}))$-IAT.

2) (Isoperimetry $\rightarrow$ Replicability): Let $\mathcal{P}$ be an N-dimensional (ρ, A)-IAT. Given $\mathcal{P}$'s membership oracle and sample access to a bounded covariance ditribution $\mathcal{D}$ over $\mathbb{R}^N$, there is an efficient $O(\rho)$-replicable algorithm that (ε, ℓ_p)-learns μ_D in $\tilde{O}\left(\frac{A^2}{N^{1-\frac{2}{p}}\varepsilon^2\rho^2}\right)$ samples.

Corollary 1.9 is somewhat weaker than its ℓ_2-analog in terms of the applicable range of ε. Namely while it is possible to derive a lower bound for ℓ_p-estimation of $\Omega(\frac{N^{1+\frac{2}{p}}}{\rho^2\varepsilon^2})$ via Corollary 1.9, the result only holds in the regime where $\varepsilon \leq \frac{1}{N^{\frac{1}{2}-\frac{1}{p}}}$. To circumvent this issue we prove a direct lower bound in the special case of the ℓ_∞-norm by an extra 'reflection' trick in our IAT analysis. This results in a near-tight characterization of replicable ℓ_∞-mean estimation:

Theorem 1.10 (Replicable ℓ_∞-Mean-Estimation): Let $\varepsilon, \rho \in (0,1)$. The ρ-replicable $(\varepsilon, \ell_\infty)$-mean-estimation problem requires

$$\tilde{\Theta}\left(\frac{N}{\varepsilon^2\rho^2}\right)$$

samples. Moreover, the lower bound holds even under Bernoulli or Gaussian distributions

Theorem 1.10 essentially resolves the complexity of the N-Coin Problem up to log factors, settling in the positive [7, Conjecture D.8]. We remark that an $\tilde{\Omega}(N)$ lower bound for N-Coins was also given in [6] under the moniker 'One-Way-Marginals' using fingerprinting. It is not clear, however, how to get the appropriate dependence on ρ and ε using their method.

3) Efficient Replicability from Relaxed Models: In the previous section we saw in the standard model, any replicable algorithm improving over the trivial union bound strategy (beyond log factors) must make progress on the efficient construction of low surface area tilings. In this section, we argue this connection can be circumvented if one is willing

to relax the model in question. We consider three relaxations that allow us to obtain efficient algorithms matching (in some cases even beating) the sample complexity implied by isoperimetric partitions: pre-processing, coordinate samples, and approximate replicability.

a) Pre-Processing: While it is true all known constructions of isoperimetric tilings have exponential time membership oracles, instead of paying this cost every time we perform a replicable procedure, we might instead hope to pay this high cost *just once* by constructing a large data structure after which membership queries can be performed in *polynomial* time. In the world of lattices, this problem is actually well-studied; it is known as the *Closest Vector Problem with Pre-processing* (CVPP). Unfortunately, existing algorithms for CVPP still run in exponential time. We show with sufficient pre-processing, it is in fact possible to solve CVPP on any lattice in *polynomial* time. More formally, we show CVPP is solvable in the *decision tree model*:

Theorem 1.11 (CVPP): Let $N \in \mathbb{N}$ and $\mathcal{L} \subset \mathbb{R}^N$. There is a depth $O(N^2 \log(N))$ decision tree $\mathcal{T}$ satisfying

1) **Pre-processing**: $\mathcal{T}$ can be constructed in $2^{\text{poly}(N)}$ time and space.
2) **Run-time**: Given $\mathcal{T}$, there is an algorithm solving CVP for all $t \in \mathbb{R}^N$ in $\text{poly}(N)$ time.

Since deterministic isoperimetric lattice tilings exist [11], all statistical upper bounds in the previous sections relying on the existence of an isoperimetric partition can in fact be executed in polynomial time after a single pre-processing cost of $2^{\text{poly}(N)}$. We remark that Theorem 1.11 may also be of independent interest. CVPP is an NP-hard problem, and prior results typically focus on improving the constants in the exponent. The decision tree model circumvents the classical hardness of CVPP by allowing access to an exponential size data structure, drawing inspiration from similar results for subset sum and other combinatorial NP hard problems [16].

b) Adaptivity and Coordinate Samples: In Section I-A2 we assumed our algorithm draws *vector* samples from an N-dimensional distribution over $\mathbb{R}^N$. In hypothesis testing (or indeed even mean estimation), sometimes the tester has more freedom and may instead choose to restrict their test to a particular *subset* of coordinates, drawing from the relevant marginal distribution. Consider, for instance, our prior example of the epidemiologist testing disease prevalence. In this setting, each 'vector sample' corresponds to a patient, and each coordinate a particular test or disease. The practitioner need not run every test on the patient (indeed this may not even be possible). Moreover, if during the procedure of the experiment some diseases are exceedingly common or rare, the practioner may wish to adaptively choose to avoid these tests and focus only on coordinates on which the result is less certain.

The equivalence of replicable mean estimation and tilings (and its corresponding lower bounds) actually holds in this coordinate sampling model as well, but only against *non-adaptive* algorithms that must choose ahead of time how many samples they'll draw for each coordinate. In the *adaptive* setting, we can actually give an efficient algorithm with coordinate sample complexity roughly $\tilde{O}(N^2)$, matching the number of coordinate samples implied by the isoperimetric lower bound for non-adaptive ℓ_∞ learning. Since the coordinate sampling model is most natural for hypothesis testing and the coin problem, we state the result in this regime:

Theorem 1.12: Let $\frac{1}{2} \geq \rho > \delta > 0$. There is a ρ-replicable algorithm solving the N-coin problem using at most $\tilde{O}\left(\frac{N^2 q_0}{(q_0-p_0)^2\rho^2}\right)$ coordinate samples and runtime.

Our algorithm requires no assumption of independence between coins. In particular, the estimates are correct and replicable even if certain diseases might be correlated.

In fact, Theorem 1.12 is really a special case of a general *adaptive composition theorem* (see Section 6 of the full paper), a computationally efficient procedure that can solve any collection of N statistical tasks replicably with $O\left(\frac{N^2}{\rho}\right)$ expected samples. The basic procedure proceeds in two steps. First, using adaptive amplification, we can solve each individual task in only $\frac{1}{\rho}$ expected samples. We then compose N such instances that are $\frac{\rho}{N}$-replicable into a ρ-replicable algorithm for the composed problem. Each of the N individual procedures costs $\frac{N}{\rho}$ samples in expectation, so linearity of expectation gives $O\left(\frac{N^2}{\rho}\right)$ total expected cost. Note that the use of average-case dependence on ρ is critical in this procedure. Composing using worst-case bounds results in a blow-up of N^3, since running each individual procedure at ρ/N-replicability costs $\frac{N^2}{\rho^2}$ samples.

c) Relaxing Replicability and the Coin Problem: Despite the above improvements, in practice sample complexity quadratic in dimension may still be prohibitively expensive. Toward this end, we consider two final relaxations of the N-coin problem where we obtain efficient algorithms with subquadratic sample complexity.

First, we consider relaxing replicability itself by allowing the output sets of the algorithm $\mathcal{A}$ between two runs to differ in at most R elements, rather than to be exactly identical.

Definition 1.13 (Approximate Replicability): Let $1 \leq R \leq N$. An algorithm $\mathcal{A}$ that outputs a set is (ρ, R)**-replicable** if for all input distributions $\mathcal{D}$,

$$\Pr_{r,S,S'}\left(|\mathcal{A}(S;r) \triangle \mathcal{A}(S';r)| \geq R\right) \leq \rho.$$

The output of the N-coin problem can be naturally viewed as a set (say the set of output large bias coins). We give an efficient adaptive (ρ, R)-replicable algorithm for the N-coin problem.

Theorem 1.14: There exists an efficient, (ρ, R)-replicable algorithm solving the N-coin problem using at most $\tilde{O}\left(\frac{q_0 N^2}{(q_0-p_0)^2 R\rho^2}\right)$ coordinate samples.

Second, we study the cost of determining only the *maximally* biased coins. Returning to our epidemiologist, while we may not have the resources to determine the prevalence of every disease, it may still be useful to determine say the 10 most prevalent, identifying a subset for which to prioritize

treatment. We design an algorithm that replicably returns a set of K coins within ε of the maximum bias $p_{\max}$.

Theorem 1.15: There is an efficient, ρ-replicable algorithm that outputs a set of at least K coins i such that $p_i \geq p_{\max} - \varepsilon$ using at most $\tilde{O}\left(\frac{N^{4/3}K^{2/3}}{\rho^2\varepsilon^2}\right)$ coordinate samples.

B. Technical Overview

We now give a high level overview of our core results and techniques, focusing on the equivalence theorem and replicability with linear overhead.

1) Replicable Algorithms and Isoperimetry:

a) Replicable Algorithms to Isoperimetric Tilings: Suppose there is a ρ-replicable algorithm $\mathcal{A}$ on m samples estimating the mean of N Bernoulli variables up to ℓ_2-error ε with probability at least $1-\delta$. We show $\mathcal{A}$ induces an approximate partition of the cube $\mathcal{C} = [1/2 - 5\varepsilon, 1/2 + 5\varepsilon]^N$ whose sets **1)** cover at least $1 - O(\rho)$ fraction of points from $\mathcal{C}$, **2)** have covering radius at most $O(\varepsilon)$, and **3)** have surface area at most $O(\rho\sqrt{m})$ (excluding the cube boundary). After scaling and translation, we obtain an approximate tiling with constant covering radius and $A \leq O(\rho\varepsilon\sqrt{m})$ surface area.

We appeal to a minimax-type argument. Consider an adversary that chooses a random mean vector $p \in \mathcal{C}$. Because $\mathcal{A}$ is correct and replicable over all biases, it must be the case that for many random strings r the deterministic procedure $\mathcal{A}(;r)$ is correct and replicable on most $p \in \mathcal{C}$. Fix such an r. For each $p \in \mathcal{C}$ on which $\mathcal{A}(;r)$ is replicable, there is some 'canonical hypothesis' $\hat{p}$ such that $\mathcal{A}(S_p, r) = \hat{p}$ with high probability when S_p is drawn from an N-Bernoulli distribution with mean p. Moreover, $\mathcal{A}(;r)$ should map any *close* biases $p, p' \in \mathcal{C}$ to the *same canonical solution* since S_p and $S_{p'}$ will be statistically indistinguishable, suggesting each $\hat{p}$ sits in a small 'bubble' of biases mapping to it. This suggests a natural candidate partitioning of the cube by these bubbles:

$$F_{\hat{p}} \coloneqq \left\{ p \in \mathcal{C} : \Pr[\mathcal{A}(;r) = \hat{p}] > \frac{3}{4} \right\}.$$

We note a similar partitioning strategy is taken in [17] to lower bound the number of random strings needed by a replicable algorithm (an orthogonal consideration to our goal of characterizing sample complexity). We discuss connections with [17] and other geometric methods in algorithmic stability in Section I-C.

Observe that by definition this partition already (nearly) satisfies Properties **(1)** and **(2)**. By replicability of $\mathcal{A}(;r)$, all but an $O(\rho)$ fraction of biases have some canonical $\hat{p}$, promising the $\{F_{\hat{p}}\}$ cover a $1 - O(\rho)$ fraction of $\mathcal{C}$. On the other hand, by correctness at most an $O(\delta)$ fraction of biases p have a canonical hypothesis $\hat{p}$ which is ε-far, meaning the sets $F_{\hat{p}}$ almost have small diameter. To ensure the sets truly have bounded diameter, we slightly modify each $F_{\hat{p}}$ by intersecting with the ε-ball $B_\varepsilon(\hat{p})$. This forces each set to have 2ε-diameter while only removing an $O(\delta)$ fraction of points from the partition.[5] Denote the new partition by $G_{\hat{p}} \coloneqq F_{\hat{p}} \cap B_\varepsilon(\hat{p})$.

Next we turn to surface area. Consider a point $p \in \partial G_{\hat{p}}$. By construction, p either lies on $\partial B_\varepsilon(\hat{p})$ or $\partial F_{\hat{p}}$. If p lies in the former, its associated canonical output $\hat{p}$ is ε-far from p, so $\mathcal{A}(;r)$ typically fails correctness on this bias. On the other hand, if p lies in the latter, there is no 'canonical hypothesis' and $\mathcal{A}(;r)$ fails replicability. The key is to observe that this is true not only of points in $\partial G_{\hat{p}}$, but for any point sufficiently nearby. Using tools from information theory, we show that for any $p, q \in \mathcal{C}$ satisfying $||p - q||_2 \leq \frac{1}{\sqrt{m}}$, $\mathcal{A}(;r)$ has similar outputs on samples from p and q. As a result, $\mathcal{A}(;r)$ fails either correctness or replicability on any point in the thickening $\partial G_{\hat{p}} + B_{\frac{1}{\sqrt{m}}}$. The desired bound now follows from considering the volume of this set. On the one hand, the volume of this thickening is roughly $\frac{1}{\sqrt{m}}$ times the surface area of $G_{\hat{p}}$.[6] On the other hand, by replicability and correctness of $\mathcal{A}(;r)$, the volume is at most $O(\rho+\delta) \leq O(\rho)$, giving the desired bound.

Finally, observe that $\mathcal{A}(;r)$ itself immediately gives a membership oracle for this approximate partition. In particular, given $p \in G_{\hat{p}}$, the oracle simply runs $\mathcal{A}(S_p; r)$ on a simulated p-biased sample S_p several times and outputs the majority. Since $\Pr[\mathcal{A}(S_p; r) = \hat{p}] \gg \frac{1}{2}$, the outcome should agree with $\hat{p}$ with high probability by Chernoff. All that is left to generate such a partition is to actually *find* a good random string r. We show most strings are good, and one can be easily found by drawing a small number and efficiently testing them for replicability.

b) Isoperimetric Tilings to Replicable Mean Estimation: Suppose we are given a (ρ, A)-IAT $\mathcal{P}$ and its associated membership oracle $\mathcal{P}(\cdot)$. We outline an oracle-efficient algorithm for replicable mean estimation for bounded covariance distributions. Our main technical contribution is an oracle efficient procedure turning any isoperimetric approximate tiling into a randomized rounding scheme such that **1)** the output after rounding is ε-close to the input with high probability, and **2)** running the rounding scheme on two inputs within distance $\varepsilon\rho\sqrt{N}/A$ leads to identical outputs with high probability when the two runs share randomness.

Given such a scheme, observe it suffices to estimate the mean non-replicably up to accuracy $\min(\varepsilon/2, \varepsilon\rho\sqrt{N}/A)$. Rounding the estimator then ensures the output is replicable and within ε distance of the true mean with high probability by the triangle inequality. For simplicity, we focus below on the regime where ε is constant; general ε error can be achieved by scaling the tiling by ε.

Given $p \in \mathcal{C}$, the most straightforward approach to rounding p would simply be to apply the membership oracle $\mathcal{P}(p)$. This

[5] Formally this means we need to assume $\delta \leq O(\rho)$. We remark that this step is not really necessary, and one can instead define an equivalence with partitions that have a 'δ-approximate diameter' of this sort. However, since $\delta \leq \rho$ is really the main regime of interest anyway, we choose to make this simplifying assumption.

[6] In reality, the volume is the integral over boundaries $\partial(G_{\hat{p}} + B_r)$ for $r \leq \frac{1}{\sqrt{m}}$. We argue there exists some r^* for which the surface area satisfies the desired bound, and take the true final partition to be $G_{\hat{p}} + B_{r^*}$, arguing this does not greatly effect the other desired properties.

clearly fails Property (2) in the worst case: no matter how close two inputs p and p' may be, as long as the segment $p - p'$ connecting them crosses a boundary of our IAT we will round to different points. This is a standard issue in replicablility (even in 1-dimension) [5]; the typical trick is to first apply a *random shift* before rounding. In our case, applying a random shift (and wrapping around $\mathcal{C}$ when necessary) ensures rounding leads to consistent outputs with probability at least $1 - \rho$ whenever the two inputs have distance at most ρ/A.

In high dimensions, however, a simple random shift is insufficient. Estimating the mean up to ρ/A accuracy requires NA^2/ρ^2 samples, so even using an isoperimetric partition ($A = \Theta(N)$) we'd require N^3 samples. The issue is we have not accounted for *direction*. Consider inputs $u^{(1)}, u^{(2)}$ that are within distance η both from each other and the boundary of the partition. Rounding $u^{(1)}$ and $u^{(2)}$ only leads to inconsistent outputs if $u^{(2)} - u^{(1)}$ points in the worst case direction, namely towards the boundary. We can avoid this by *randomly rotating* our input before shifting it. The resulting difference vector $u^{(2)} - u^{(1)}$ then points in a random direction and a simple calculation shows the worst-case direction has size $\frac{1}{\sqrt{N}} ||u^{(2)} - u^{(1)}||_2$ in expectation. This saves a $\sqrt{N}$ factor, meaning our original points only need to be within distance $O(\rho\sqrt{N}/A)$ as desired.

2) Lower Bounds for the N-Coin Problem: Recall for ℓ_∞-estimation and the N-Coin Problem, the procedure described above only gives a tight sample lower bound of $\Omega(N\varepsilon^{-2}\rho^{-2})$ vector samples when $\varepsilon \leq \frac{1}{\sqrt{N}}$. We now discuss how to modify the argument to give a tight bound in all regimes. For convenience we work directly with the N-Coin Problem, and assume that the algorithm invokes m flips for each of the coin (alternatively, the algorithm takes m vector samples).

Similar to the argument in the ℓ_2-case, we look at the set of possible canonical outputs $\hat{o} \in \{\textsc{Accept}, \textsc{Reject}\}^N$, and the approximate partition $\{F_{\hat{o}}\}$ over $\mathcal{C} = \left[\frac{1}{4}, \frac{3}{4}\right]^N$ induced by the algorithm in the same manner. If we could show the surface area of the boundaries of $\{F_{\hat{o}}\}$ (excluding the cell boundary) is at least $\Omega(\sqrt{N}\varepsilon^{-1})$, we would be able to use a similar argument to the ℓ_2 case to show the fraction of non-replicable points is at least $\sqrt{N/m}\varepsilon^{-1} \leq O(\rho)$, implying the desired sample complexity lower bound on m.

The main difficulty in the ℓ_∞ setting is an issue we brushed under the rug in the previous section: the cube boundary. In particular, the naive way of lower bounding the surface area of $\{F_{\hat{o}}\}$ is to apply the isoperimetric inequality to $\partial(\cup_{\hat{o}} F_{\hat{o}} \cap \mathcal{C})$, then subtract out the boundary of the cube. Since we now measure error in ℓ_∞-norm, however, we can only bound the radius of $F_{\hat{o}}$ by $\sqrt{N}\varepsilon$ and the above method gives surface area $A \geq \sqrt{N}\varepsilon^{-1} - O(N)$, useless when $\varepsilon > \frac{1}{\sqrt{N}}$.

To circumvent this, we need to somehow apply the isoperimetric inequality to $\partial F_{\hat{o}} \backslash \partial\mathcal{C}$ directly. To this end, first observe that, by correctness, $F_{\hat{o}}$ can only intersect a δ-fraction of faces of $\mathcal{C}$ not incident to the corner $\hat{o}$. Moreover, if $F_{\hat{o}}$ only intersects such faces, we can create a valid surface by *reflecting* $F_{\hat{o}}$ across the cube boundary. This forces points on the cube boundary to become interior while otherwise 'copying' the boundary of $F_{\hat{o}}$ itself 2^N times. Since reflecting only changes the ℓ_∞-radius by a constant factor, we can now apply the isoperimetric inequality to the reflected set with no asymptotic loss to get the desired lower bound on $\partial F_{\hat{o}}$.

3) Adaptivity: Prior replicable algorithms in the literature are *non-adaptive*: they draw a fixed number of samples ahead of time, typically incurring a quadratic dependence on ρ as a result. We show this strategy is wasteful. By instead allowing the algorithm to terminate early based on initial observations, we can reduce this cost to just *linear* expected overhead. As discussed in Section I-A3, this also leads to an adaptive composition theorem with improved overhead.

Here we overview our most basic adaptive algorithm, testing the bias of a single coin (say between $1/3$ and $2/3$). Prior algorithms based on statistical queries compute the empirical bias of the coin using a *fixed* number of samples and compare it with a random threshold, ensuring sufficient samples are drawn such that even if the threshold is within $O(\rho)$ of the bias p, our estimate still lands on the correct side. Our adaptive algorithm samples a random threshold r and draws samples adaptively until it determines whether the true bias p lies above or below the threshold r. The key observation is that when the true bias p and the random threshold r are far apart, we only need $\frac{1}{|r-p|^2}$ samples to determine with high confidence whether the true bias is above or below the threshold. Since r uniformly random, $|r - p|$ is (roughly) uniform over $(\rho, 1/3)$ and the expected sample complexity is

$$\int_\rho^{1/3} \frac{1}{x^2} dx = O\left(\frac{1}{\rho}\right).$$

Using similar ideas, we also build an adaptive algorithm for the heavy hitters problem. This allows us to run an adaptive variant of replicability amplification (similar to [5], [6]) to show *any* replicable algorithm can be run with only linear expected overhead.

C. Further Related Work

a) Replicability: Algorithmic replicability was independently introduced in [5], [18]. Replicable algorithms have since been developed for PAC Learning [6], [19], reinforcement learning [7], [20], bandits [21], [22], clustering [23], and large-margin halfspaces [5], [24]. Several works have shown tight statistical connections between replicability and other notions of algorithmic stability [6], [17], [19], [25]–[27]. Most closely related to our work are the discussed algorithms (and lower bounds) for N-Coins and mean estimation problems in [6], [7] respectively, and the work of [17] studying 'list' or 'certificate' replicability for N-Coins. The latter in particular uses a similar partitioning strategy to our lower bound, but relies on totally different properties of the partition.

b) Geometry and Algorithmic Stability: Our work adds to a growing line of connections between geometry, topology, and algorithmic stability. Such ideas were first introduced in the study of pure differential privacy in [28], where packing lower bounds are now a standard tool [29]–[32]. Impossibility results for related notions of replicability, specifically list

replicability, certificate replicability, and global stability have been obtained via geometric and topological tools [17], [25], [26], in particular via the Sperner lemma and variants of Borsuk-Ulam.

c) Tilings and Rounding: The basic connection between replicability, tilings, and randomized rounding was first observed in [5]. The authors used 1-dimensional randomized rounding to give the first replicable algorithms for statistical queries and heavy hitters, and the high dimensional scheme of [13] to build a replicable PAC-learner for large margin halfspaces. The authors also analyzed rounding via cubical tiling, equivalent to independent handling of each coordinate.

There are many known constructions of isoperimetric tilings [9]–[11], [13], [14]. Our work is mostly closely related to [13], who also observe their construction induces a 'noise resistant' rounding scheme. Both our work and [13] critically rely on the Buffon needle theorem to analyze surface area and noise resistence. The main difference is that [13] study a specific randomized framework that in some sense 'automatically' results in rounding, while we show how to take an arbitrary tiling and transform it into a rounding algorithm.

D. Full Version of This Work

We refer the reader to the full version of this work for more and rigorous discussions of our results [33].

Acknowledgment

The authors would like to thank Jelena Bradic, Byron Chin, Nicholas Genise, Rex Lei, Shachar Lovett, Toniann Pitassi, Mark Schultz, and Saket Shah for many helpful discussions and thoughtful comments.

References

[1] C. G. Begley and L. M. Ellis, "Raise standards for preclinical cancer research," *Nature*, vol. 483, no. 7391, pp. 531–533, 2012.

[2] J. P. Ioannidis, "Contradicted and initially stronger effects in highly cited clinical research," *Jama*, vol. 294, no. 2, pp. 218–228, 2005.

[3] M. Baker, "1,500 scientists lift the lid on reproducibility," *Nature*, vol. 533, no. 7604, 2016.

[4] J. P. Ioannidis, "Why most published research findings are false," *PLoS medicine*, vol. 2, no. 8, p. e124, 2005.

[5] R. Impagliazzo, R. Lei, T. Pitassi, and J. Sorrell, "Reproducibility in learning," in *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pp. 818–831, ACM, 2022.

[6] M. Bun, M. Gaboardi, M. Hopkins, R. Impagliazzo, R. Lei, T. Pitassi, S. Sivakumar, and J. Sorrell, "Stability is stable: Connections between replicability, privacy, and adaptive generalization," in *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC*, pp. 520–527, ACM, 2023.

[7] A. Karbasi, G. Velegkas, L. F. Yang, and F. Zhou, "Replicability in reinforcement learning," *CoRR*, vol. abs/2305.19562, 2023.

[8] T. C. Hales, "The honeycomb conjecture," *Discrete & computational geometry*, vol. 25, pp. 1–22, 2001.

[9] C. A. Rogers, "A note on coverings and packings," *Journal of the London Mathematical Society*, vol. s1-25, no. 4, pp. 327–331, 1950.

[10] G. Butler, "Simultaneous packing and covering in euclidean space," *Proceedings of the London Mathematical Society*, vol. 3, no. 4, pp. 721–735, 1972.

[11] D. Micciancio, "Almost perfect lattices, the covering radius problem, and applications to ajtai's connection factor," *SIAM Journal on Computing*, vol. 34, no. 1, pp. 118–169, 2004.

[12] E. Mook and C. Peikert, "Lattice (list) decoding near minkowski's inequality," *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 863–870, 2021.

[13] G. Kindler, A. Rao, R. O'Donnell, and A. Wigderson, "Spherical cubes: optimal foams from computational hardness amplification," *Communications of the ACM*, vol. 55, no. 10, pp. 90–97, 2012.

[14] A. Naor and O. Regev, "An integer parallelotope with small surface area," *Journal of Functional Analysis*, vol. 285, no. 10, p. 110122, 2023.

[15] T. Baigneres, P. Junod, and S. Vaudenay, "How far can we go beyond linear cryptanalysis?," in *International Conference on the Theory and Application of Cryptology and Information Security*, pp. 432–450, Springer, 2004.

[16] F. Meyer auf der Heide, "A polynomial linear search algorithm for the n-dimensional knapsack problem," *Journal of the ACM (JACM)*, vol. 31, no. 3, pp. 668–676, 1984.

[17] P. Dixon, A. Pavan, J. Vander Woude, and N. Vinodchandran, "List and certificate complexities in replicable learning," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[18] B. Ghazi, R. Kumar, and P. Manurangsi, "User-level differentially private learning via correlated sampling," in *Advances in Neural Information Processing Systems 34*, pp. 20172–20184, 2021.

[19] A. Kalavasis, A. Karbasi, S. Moran, and G. Velegkas, "Statistical indistinguishability of learning algorithms," in *International Conference on Machine Learning, ICML*, 2023.

[20] E. Eaton, M. Hussing, M. Kearns, and J. Sorrell, "Replicable reinforcement learning," *arXiv preprint arXiv:2305.15284*, 2023.

[21] H. Esfandiari, A. Kalavasis, A. Karbasi, A. Krause, V. Mirrokni, and G. Velegkas, "Replicable bandits," in *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*, OpenReview.net, 2023.

[22] J. Komiyama, S. Ito, Y. Yoshida, and S. Koshino, "Improved algorithms for replicable bandits," 2023.

[23] H. Esfandiari, A. Karbasi, V. Mirrokni, G. Velegkas, and F. Zhou, "Replicable clustering," *CoRR*, vol. abs/2302.10359, 2023.

[24] A. Kalavasis, A. Karbasi, K. G. Larsen, G. Velegkas, and F. Zhou, "Replicable learning of large-margin halfspaces," *arXiv preprint arXiv:2402.13857*, 2024.

[25] Z. Chase, B. Chornomaz, S. Moran, and A. Yehudayoff, "Local borsuk-ulam, stability, and replicability," *CoRR*, vol. abs/2311.01599, 2023.

[26] Z. Chase, S. Moran, and A. Yehudayoff, "Replicability and stability in learning," *CoRR*, vol. abs/2304.03757, 2023.

[27] S. Moran, H. Schefler, and J. Shafer, "The bayesian stability zoo," *CoRR*, vol. abs/2310.18428, 2023.

[28] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 705–714, 2010.

[29] A. Bhaskara, D. Dadush, R. Krishnaswamy, and K. Talwar, "Unconditional differentially private mechanisms for linear queries," in *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pp. 1269–1284, 2012.

[30] A. Nikolov, K. Talwar, and L. Zhang, "The geometry of differential privacy: the sparse and approximate cases," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 351–360, 2013.

[31] A. Beimel, H. Brenner, S. P. Kasiviswanathan, and K. Nissim, "Bounds on the sample complexity for private learning and private data release," *Machine learning*, vol. 94, pp. 401–437, 2014.

[32] S. Vadhan, "The complexity of differential privacy," *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich*, pp. 347–450, 2017.

[33] M. Hopkins, R. Impagliazzo, D. Kane, S. Liu, and C. Ye, "Replicability in high dimensional statistics," *arXiv preprint arXiv:2406.02628*, 2024.