Robust Traceability from Trace Amounts[*]

Cynthia Dwork[†] Adam Smith[‡] Thomas Steinke[§]
Jonathan Ullman[¶] Salil Vadhan[‖]

January 8, 2018

Abstract

The privacy risks inherent in the release of a large number of summary statistics were illustrated by Homer *et al.* (*PLoS Genetics*, 2008), who considered the case of one-way marginals of SNP allele frequencies obtained in a genome-wide association study: Given a large number of minor allele frequencies from a case group of individuals diagnosed with a particular disease, together with the genomic data of a single target individual and statistics from a sizable reference dataset independently drawn from the same population, an attacker can determine with high confidence whether or not the target is in the case group.

In this work we describe and analyze a simple attack that succeeds even if the summary statistics are significantly distorted, whether due to measurement error or noise intentionally introduced to protect privacy. Our attack only requires that the vector of distorted summary statistics is close to the vector of true marginals in ℓ_1 norm. Moreover, the reference pool required by previous attacks can be replaced by *a single sample* drawn from the underlying population.

The new attack, which is not specific to genomics and which handles Gaussian as well as Bernouilli data, significantly generalizes recent lower bounds on the noise needed to ensure differential privacy (Bun, Ullman, and Vadhan, STOC 2014; Steinke and Ullman, JPC 2017), obviating the need for the attacker to control the exact distribution of the data.

[*] A preliminary version of this work appeared in *FOCS 2015*.

[†] Harvard University. Part of this work was done while at Microsoft and while visiting the Simons Institute for the Theory of Computing. Cynthia Dwork was supported by the Simons Foundation and by the DIMACS/Simons Collaboration in Cryptography through NSF grant CNS-1523467. dwork@seas.harvard.edu

[‡] Boston University. Part of this work was done while at Pennsylvania State University and while visiting Harvard University's Center for Research on Computation & Society, where he was supported by a Simons Investigator award to Salil Vadhan. Adam Smith was supported by NSF award IIS-1447700 and a Google Faculty Award. ads22@bu.edu

[§] IBM Research – Almaden. Part of this work was done while at Harvard University. Thomas Steinke was supported by NSF grants CNS-1237235, CCF-1116616, and CCF-1420938. tracing@thomas-steinke.net

[¶] Northeastern University. Jonathan Ullman was supported by a Junior Fellowship from the Simons Society of Fellows. Part of this work was done while the author was a postdoctoral fellow at Columbia University. jullman@ccs.neu.edu

[‖] Harvard University. Salil Vadhan was supported by NSF grant CNS-1237235, a gift from Google, Inc., and a Simons Investigator grant. salil_vadhan@harvard.edu

Contents

1 Introduction

Given a collection of (approximate) summary statistics about a dataset, and the precise data of a single target individual, under what conditions is it possible to determine whether or not the target is a member of the dataset? This *tracing* problem is the focus of our work.

Questions of this type arise in many natural situations in which membership in the dataset is considered sensitive; indeed, this is typically the reason for choosing to publish summary statistics, as opposed to releasing the raw data. In a scenario that is prominent in the literature, the dataset contains genomic information about a *case group* of individuals with a specific medical diagnosis, as in a genome-wide association study (GWAS), and the summary statistics are SNP allele frequencies, *i.e. one-way marginals*. Specifically, if each person's data consists of d binary attributes, we consider a mechanism that releases (an approximation to) the average value of the each of the d attributes. Homer *et al.* [HSR+08] demonstrated the privacy risks inherent in this scenario, presenting and analyzing a tracing algorithm for membership in a GWAS case group, provided the attacker also has access to allele frequencies for a reference group of similar ancestral make-up as that of the case group.

It came as a surprise to the genomics research community that the trace amount of DNA contributed by an individual is enough to determine membership in the case group with high statistical confidence. The result had a major practical impact in the form of very restrictive policies governing access to allele frequency statistics in studies funded by the US National Institutes of Health and the Wellcome Trust. Follow-up analytical works provide alternative tests and asymptotic analyses of tradeoffs between the size of the test set, the size of a reference dataset, power, confidence, and number of measurements [SOJH09].

As in the follow-up works, the analysis in Homer *et al.* assumes that *exact* statistics are released, leaving open the possibility that the attack may be foiled if the statistics are distorted, for example, due to measurement error (which can be highly correlated across the statistics), or because noise is intentionally introduced in order to protect privacy.

In this paper, we show that one can test if an individual is present in the case group even when the one-way marginals are considerably distorted before being released. We give a single tracing attack that applies to *all* mechanisms that produce sufficiently accurate estimates of the statistics in question, rather than to just the single mechanism that outputs exact statistics.

A line of work initiated by Dinur and Nissim [DN03] provides attacks of this flavor for certain kinds of statistics, showing that all mechanisms that release "too many" answers that are "too accurate" are subject to devastating "reconstruction attacks," which allow an adversary to determine the private data of almost all individuals in a dataset. These attacks, which immediately give lower bounds on noise needed to avoid blatant non-privacy, have been extended in numerous works [DMT07, DY08, KRSU10, De12, KRS13, MN12, FMN13, NTZ13].

These reconstruction attacks do not generally apply in the setting of Homer *et al.*, since they either require that the amount of noise introduced for privacy is very small (less than the sampling error), or require an exponential number of statistics, or do not apply to statistics that are as simple (namely, attribute frequencies), or require that the adversary have a significant amount of auxiliary information about the other individuals in the dataset.

Of course, complete reconstruction is an extreme privacy failure: the privacy of essentially every member of the dataset is lost! Conversely, protection from complete reconstruction is a very low barrier for a privacy mechanism. What if we are more demanding, and ask that an attacker not be able to determine whether an individual is present or absent from the dataset, that is, to *trace*? This in/out protection is the essence of differential privacy, and the question of how much noise is needed to ensure differential privacy, first studied in [HT10], has seen many recent developments [Ull13, BUV14, DNT14, HU14, SU14, SU17]. By shifting the goal from reconstructing to tracing, these works obtain lower bounds on noise for settings where reconstruction is impossible.

In particular, the papers [BUV14, SU17] provide tracing attacks, based on the use of *fingerprinting codes* [BS98, Tar08], that operate given attribute frequencies of the database with only non-trivial accuracy. However, they require that the attribute frequencies of the underlying population are drawn from a particular, somewhat unnatural distribution, and that the attacker has very accurate knowledge of these frequencies. We remark that such knowledge is the "moral equivalent," in this literature, to having a large reference population, in the genomics literature.

In this paper, we generalize the attacks based on fingerprinting codes in several ways to considerably broaden their applicability:

- The population's attribute frequencies can be drawn from any distribution on $[0,1]$ that is sufficiently smooth and spread out, including, for example, the uniform distribution on $[0,1]$ or a large subinterval. The tracing algorithm does not depend on the distribution.
- Instead of knowing the population attribute frequencies, it suffices for the attacker to have a *single* reference sample from the population.
- We show that similar attacks can be applied to Gaussian data (rather than binary data) for mechanisms that release too many attribute averages with nontrivial accuracy.

Our results provide a common generalization of the fingerprinting results and the results of Homer et al, showing they are special cases of a much broader phenomenon.

Like the fingerprinting attacks of [BUV14, SU17], the lower bounds on noise implied by our attacks nearly match the upper bounds on noise sufficient to ensure the strong guarantees of differential privacy, for example, via the Gaussian or Laplace mechanisms [DN03, BDMN05, DMNS06, DKM+06, DRV10]). Thus, the cost in utility for avoiding our attacks is nearly the same as the cost for avoiding the much larger class of attacks that differential privacy prevents, where the dataset can be arbitrary and the attacker can know everything about it, except whether or not the target individual is present in the dataset.

1.1 Model and Assumptions

Distributional Assumption. The database consists of n independent samples from a *population*, which is given by a product distribution $\mathcal{P}_p$ on $\{\pm 1\}^d$. The vector $p \in [-1,1]^d$ specifies the expectation of a sample from $\mathcal{P}_p$. That is, to sample $x \sim \mathcal{P}_p$, we set $x_j = 1$ with probability $(1+p_j)/2$ and set $x_j = -1$ with probability $(1-p_j)/2$, independently for each j.

The vector p represents unknown statistics about the population; p is unknown to both the mechanism and the privacy attacker.[1] The vector p is itself drawn from the product distribution $\mathcal{D}$ on $[-1,1]^d$ with the j^{th} marginal having probability density function $\rho_j : [-1,1] \to \mathbb{R}$. In the case of genomics, we can think of the distribution $\mathcal{D}$ as capturing, for example, differences between populations (although of course in reality this would not be a product distribution). Our attacks will succeed even if the mechanism knows $\mathcal{D}$ but the attacker does not, provided each ρ_j is sufficiently smooth and spread out *e.g.*, if ρ_j is uniform on a large enough subinterval of $[0,1]$).

[1] If the mechanism knows p then the problem becomes vacuous: it could simply ignore the data and publish p.

Accuracy of the Mechanism. The (possibly randomized) *mechanism* $\mathcal{M}$ receives n independent samples $x_1, \cdots, x_n \in \{\pm 1\}^d$ drawn from $\mathcal{P}_p$ (after p is initially drawn from $\mathcal{D}$), and outputs a vector $q \in [-1,1]^d$ with $q \approx \bar{x} = \frac{1}{n}\sum_{i\in[n]} x_i \approx p$. That is, $\mathcal{M}$ provides approximate one-way marginals. We say $\mathcal{M}$ is *α-accurate* if for all $j \in [d]$ we have $\left|\mathbb{E}\left[q^j\right] - p^j\right| \leq \alpha$ for all possible values of p, where the expectation is taken over the randomness of $\mathcal{M}$ and the sample x. We require this to hold even when we condition on $x^{j'}$ and $q^{j'}$ for $j' \neq j$. This is a very weak accuracy requirement, as it only refers to the *bias* of the statistics, namely $\mathbb{E}[q] - p$. We also require that q is bounded in $[-1,1]^d$, so if the mechanism adds unbounded noise, we should truncate the answers, which may increase the bias.

The Attacker. The *privacy attacker* $\mathcal{A}$ receives two samples in $\{\pm 1\}^d$, the target y and the reference z, where z is drawn independently from the population $\mathcal{P}_p$, together with the output q of $\mathcal{M}$ on a dataset $x_1, \ldots, x_n$, and produces an answer, either IN or OUT. The attacker's answer indicates whether or not it believes y is among the $x_1, \cdots, x_n$ given to $\mathcal{M}$. The attacker is guaranteed that the reference sample z is drawn from $\mathcal{P}_p$ independent from everything else. The attacker must satisfy two properties:

- *Soundness:* If y is drawn from $\mathcal{P}_p$ independent from the view of $\mathcal{M}$ (i.e. independent from q), then $\mathcal{A}$ should output IN with probability at most s.
- *Completeness:* Choose i uniformly from $[n]$ and set $y = x_i$. Then $\mathcal{A}$ should output IN with probability at least c. The probability is over all the random choices: i, x, z, and the coin flips of $\mathcal{A}$ and $\mathcal{M}$.

These conditions are interesting when $c \gg s$, as when $c \leq s$ they are trivially satisfied by having $\mathcal{A}$ always output IN with probability c. To interpret this, think of y as the data of a member of the population and $\mathcal{A}$ wants to determine whether or not y is in the dataset (case group) given to $\mathcal{M}$. For $\mathcal{A}$ to be considered successful we require that it can identify a random member of the dataset with reasonably high probability (given by the completeness parameter c), whilst, if y is not in the dataset, it is erroneously claimed otherwise with negligible probability (given by the (un)soundness parameter s). The reference sample z is some minimal auxiliary information about the population that $\mathcal{A}$ can use.

1.2 Our Results

Theorem 1 (Main – Informal)**.** *There is a universal constant $\alpha > 0$ such that for every $\delta > 0$, $n \in \mathbb{N}$, and $d \geq O(n^2 \log(1/\delta))$, there exists an attacker $\mathcal{A} : \{\pm 1\}^d \times [-1,1]^d \times \{\pm 1\}^d \to \{\mathrm{IN}, \mathrm{OUT}\}$ the following holds.*

Let $\mathcal{D}$ be a product distribution on $[-1,1]^d$ such that each marginal satisfies a technical smoothness condition (Definitions 5 and 26). Let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ be α-accurate. Let $p \sim \mathcal{D}$ and $x_1, \cdots, x_n, y, z \sim \mathcal{P}_p$. Let $q \sim \mathcal{M}(x_1, \cdots, x_n)$. Then

$$\mathbb{P}[\mathcal{A}(y,q,z) = \mathrm{IN}] \leq \delta \quad \textit{and} \quad \mathbb{P}[\exists\ i \in [n]\ \ \mathcal{A}(x_i,q,z) = \mathrm{IN}] \geq 1 - \delta.$$

Thus, if the first input (y) to $\mathcal{A}$ is a random independent element of the population, then $\mathcal{A}$ will accept with probability at most $s \leq \delta$ (the probability space includes the selection of y), but if the first input is a random element of the dataset (x_i for a random i), then $\mathcal{A}$ will accept with probability at least $c \geq (1-\delta)/n$. Thus, the result is nontrivial when $\delta < (1-\delta)/n$ (e.g. $\delta = o(1/n)$).

We discuss a number of features and extensions of the result.

Dimensionality Needed. The dimensionality d of the data needed for the attack is $d = \tilde{O}(n^2)$ for $\delta = 1/2n$, which is tight up to polylogarithmic factors for achieving constant accuracy α. Indeed, it is possible to answer $d = \tilde{\Omega}(n^2)$ one-way marginals with accuracy $\alpha = o(1)$, while satisfying the strong guarantee of $(o(1), 1/n^{\omega(1)})$-differential privacy [DN03, BDMN05, DKM+06, DMNS06, DRV10].[2] (Our attack implies that no mechanism satisfying the above conditions can be $(0.1, 1/4n)$ differentially private.) For the one-way marginals we consider, the number of statistics released equals the dimensionality d of the data, but for richer families of statistics, the dimensionality is the more significant parameter. Indeed, many more than n^2 statistics can be released if the dimensionality d of the data is smaller than n^2—the algorithms of [BLR08, HR10, RR10, DRV10] can release a number of statistics that is nearly exponential in $n/\sqrt{d}$.

Beyond the $d = \Theta(n^2)$ Barrier. The price for our very weak assumptions – weakly accurate answers and only a single reference sample – is that we (provably) need $d = \Omega(n^2)$ and can only trace a single individual. With more accurate answers and a larger reference pool, a slightly modified version of our attacker can trace with smaller d, and can trace many individuals in the dataset: if the mechanism is α-accurate (for some $\alpha \geq n^{-1/2}$), and we are given roughly $1/\alpha^2$ independent reference samples from the distribution, then we trace when the dataset has dimension only $O(\alpha^2 n^2)$. Moreover, we can successfully trace $\Omega(1/\alpha^2)$ individuals in the dataset, yielding a completeness probability of $c = \Omega(1/\alpha^2 n)$ (Section 3).

Weaker Soundness Conditions. The soundness of our attack does not rely on any properties of the distribution $\mathcal{D}$, the accuracy of $\mathcal{M}$, the relation between d, n, and δ, or even the distribution of the rows $x_1, \ldots, x_n$. It only requires that conditioned on q, the individuals y and z are sampled independently from the same product distribution. Thus, the attack can be carried out under only the latter assumption, and if it says IN, one can safely conclude $y \in \{x_1, \ldots, x_n\}$.

Higher-Power Attacks. Our completeness probability of $c = \Theta(1/\alpha^2 n)$ is essentially tight, as a mechanism $\mathcal{M}$ that outputs the averages on a subsample of size $O(1/\alpha^2)$ will be accurate but only allows tracing at most an $O(1/\alpha^2 n)$ fraction of individuals in the dataset

However, if we assume that $\mathcal{M}$ is *symmetric*, then we can get around this. That is, if we assume that $\mathcal{M}$ can be written as $\mathcal{M}(x_1, \cdots, x_n) = \mathcal{M}'(\overline{x})$ (where $\overline{x} = \frac{1}{n}\sum_{i \in [n]} x_i \in [-1,1]^d$ is the average of the sample), then we can prove that

$$\forall i \in [n] \quad \mathbb{P}[\mathcal{A}(x_i, q, z) = \text{IN}] \geq 1 - \delta.$$

Note that with this high-power guarantee ($c \geq 1 - \delta$), it is meaningful to take δ to be a fixed constant (e.g. the standard significance level of .05).

[2] An algorithm that operates on datasets is (ε, δ)-differentially private if for all datasets S, S' differing in the data of a single individual and every event E, the probability of E when the dataset is S is at most δ plus e^ε times the probability of E when the dataset is S'.

The Distribution $\mathcal{D}$. As noted above, we impose a technical regularity condition on the distribution $\mathcal{D}$, requiring that its marginals ρ_j are sufficiently smooth and spread out. This includes distributions such as the uniform distribution on a large subinterval and the family of Beta distributions.

Some assumptions on $\mathcal{D}$ are necessary. For example, if each marginal ρ_j were supported on a subinterval of length at most α, then the mechanism could give accurate answers by just producing a vector $q \in [-1,1]^d$ in the support of $\mathcal{D}$ and not using the dataset at all. This shows that the ρ_j need to be sufficiently "spread out". To see why "smoothness" is necessary, suppose that ρ_j were concentrated on two points p^* and p^{**} that are reasonably far apart (farther than 2α). Then the mechansim can simply test whether the average of the data elements exceeds $(p^* + p^{**})/2$ and, if so, output $\max\{p^*, p^{**}\}$; otherwise output $\min\{p^*, p^{**}\}$. While this mechanism is not differentially private (a guarantee against tracing in the worst case), with high probability over the choice of the dataset this mechanism is insensitive to small changes in the dataset, *i.e.*, changing one row will not change the output. This makes tracing impossible.

Real-Valued Data. In many settings, the database takes values in $\mathbb{R}^{n\times d}$ rather than $\{\pm 1\}^{n\times d}$. We show that, if the data $x_1, \cdots, x_n$ are independent samples from a multivariate Gaussian (with no covariances), the same attack can be carried out. We require an upper bound $\sigma^2_{\max}$ on the variance of the data entries and assume that the coordinate means are again drawn from a smooth and spread out distribution. In this setting we require $d = O(n^2 \sigma^2_{\max} \log(1/\delta))$.

1.3 Description of The Attack

Like the attacks in previous tracing work for the genomic setting [HSR+08, SOJH09, BRS+09, JYW+09, ZPL+11] and in the fingerprinting setting [Tar08, SU14], our attack uses a simple scoring function to make its decision. The scoring function works incrementally, with each marginal (SNP) making a separate contribution. The attack is described in full in Figure 1.

```
$\mathcal{A}_{\delta,d}(y,q,z)$

1. Input:  $y, z \in \{\pm 1\}^d$ and $q \in [-1,1]^d$.
2. Compute $\langle y,q \rangle = \sum_{j\in[d]} y^j \cdot q^j$ and $\langle z,q \rangle = \sum_{j\in[d]} z^j \cdot q^j$.
3. If $\langle y,q \rangle - \langle z,q \rangle > \tau := \sqrt{8d\ln(1/\delta)}$, output IN; otherwise output OUT.
```

Figure 1: Our Privacy Attack

The key features of the adversary are that it only sees the data of the user y being traced, plus a reference sample z (in addition, of course, to seeing the output q), and does not depend on the mechanism $\mathcal{M}$, the unknown mean p, or the distribution $\mathcal{D}$ on p.

1.4 Comparison with Previous Work

As mentioned above, our model and results provide a common generalization of lines of work from several fields.

- Work in the genomics community [HSR+08, BRS+09, VH09, SOJH09, JYW+09] has so far focused on the case where *exact* statistics are available to the attacker ($\alpha = 0$ in our

formalism). With a reference sample of $\Omega(n)$ individuals, they showed that $d = \Theta(n)$ attributes are necessary and sufficient, while with a constant-sized reference pool, $d = \Theta(n^2)$ is required [SOJH09]). Our first attack uses $\Theta(n^2 \cdot \log n)$ statistics with a reference pool of size 1, and makes only a minimal accuracy assumption (a constant bound α on the bias).
Our second attack requires only $d = \tilde{O}(\alpha^2 n^2)$ statistics if the mechanism is α-accurate (for some $\alpha \geq n^{-1/2}$) and the reference pool is of size $O(\log(n)/\alpha^2)$, in which case it can also successfully trace $\Omega(1/\alpha^2)$ individuals in the dataset.
Im *et al.* [IGNC12] use (exact) regression coefficients instead of marginals as the basis of an attack, with similar results to the case of marginals.
- Work on fingerprinting attacks [BUV14, SU17] corresponds to our setting of a constant α, but assumes that p is drawn from a specific distribution $\mathcal{D}$, and the attacker $\mathcal{A}$ knows p exactly (essentially, an infinite reference pool). The dimensions required in their attacks are similar to ours ($d = \Theta(n^2)$).

We note that previous work has focused on categorical data, but our results extend to the setting of normally-distributed real-valued data.

Other Work on Genetic Privacy. The literature contains attacks based on various types of published aggregate statistics, *e.g.*, allele frequencies, genetic frequencies, and various quantitive phenotypes such as cholesterol levels [HSR$^+$08, JYW$^+$09, WLW$^+$09, IGNC12]; see [EN14] for a survey. Particularly exciting (or troubling) is the work of Wang *et al.* [WLW$^+$09] that exploits correlations among different SNPs. Not only do their attacks require relatively few SNPs, but they go beyond in/out privacy compromise, actually reconstructing SNPs of members of the case group. In our view, the message of these works and ours, taken as a whole, is that information combines in surprising ways, aggregation should not be assumed to provide privacy on its own, and rigorous approaches to controlling privacy risk are *necessary*.

2 Tracing with a Single Reference Sample

Now we analyze our attack (given in Figure 1) and thereby prove Theorem 1.

2.1 Soundness Analysis

Proposition 2 (Soundness)**.** *Let $q, p \in [-1,1]^d$. Suppose $y, z \sim \mathcal{P}_p$ are independent from each other and from q. Then*

$$\mathbb{P}[\mathcal{A}_{\delta,d}(y,q,z) = \text{IN}] \leq \delta.$$

Proof. We can view p and q as fixed. Since y and z are identically distributed, $\mathbb{E}[\langle y,q\rangle - \langle z,q\rangle] = 0$. Since y and z are independent samples from a product distribution, we have that $\langle y,q\rangle - \langle z,q\rangle = \sum_{i\in[d]}(y^j - z^j)\cdot q^j$ is the sum of $2d$ independent random variables each of which is bounded by $\max\{\left\|y\right\|_\infty, \|z\|_\infty\} \cdot \|q\|_\infty \leq 1$. Thus, by Hoeffding's inequality,

$$\mathbb{P}[\langle y,q\rangle - \langle z,q\rangle > \tau] \leq e^{-\tau^2/4d} = \delta,$$

as required. □

Remark 3. *Proposition 2 makes no assumptions about q. Thus soundness holds even if $\mathcal{M}$ is not accurate or if y, z are not sampled from the true population – they need only be sampled from the same product distribution.*

2.2 Correlation Analysis

To prove completeness we must show that $\langle x_i, q\rangle - \langle z, q\rangle > \tau$ with good probability for a random $i \in [n]$ when the mechanism's output is α-accurate. First we give a formal definition of accuracy:

Definition 4 (Accuracy)**.** *We say $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ is α-accurate if*

$$\|M(x) - \overline{x}\|_\infty \le \alpha$$

for all $x \in \{\pm 1\}^{n\times d}$, where $\overline{x} \in [-1,1]^d$ is the average of the rows of x.

Note that, for simplicity, we assume the accuracy condition holds with probability 1. However, a high probability bound would suffice, but we would need to carry an extra parameter through our analysis.

In Section 4.1, we discuss mechanisms that satisfy a weaker "ℓ_1" accuracy condition.

We begin by showing that, under our regularity assumption on $\mathcal{D}$,

$$\mathbb{E}\left[\sum_{i\in[n]} (\langle x_i, q\rangle - \langle z, q\rangle)\right] \ge Cn\tau$$

for an appropriate constant $C > 1$.

Intuitively, $\sum_{i\in[n]}\langle x_i, q\rangle$ measures how much the output $q \in [-1,1]^d$ of $\mathcal{M}$ correlates with the input $x_1, \cdots, x_n \in \{\pm 1\}^d$ of $\mathcal{M}$, whereas $\langle z, q\rangle$ measures how much a random member of the population correlates with q. Thus we are proving that the output of $\mathcal{M}$ is more correlated with the input of $\mathcal{M}$ than with an independent sample from the population.

By linearity of expectations it suffices to show that $\mathbb{E}\left[\sum_{i\in[n]} x_i^j q^j - z^j q^j\right] \ge Cn\tau/d$ for each $j \in [d]$. We now focus on a fixed $j \in [d]$ and, for clarity, omit the superscript.

First some notation: Let $p \sim \rho$ denote that $p \in \mathbb{R}$ is drawn according to the probability distribution given by ρ (e.g. ρ is a probability density function $\rho : \mathbb{R} \to \mathbb{R}$). For $p \in [-1,1]$, let $x \sim p$ denote that $x \in \{\pm 1\}$ is drawn with $\mathbb{E}[x] = p$. Let $x_{1\cdots n} \sim \rho$ denote that $x_1, \cdots x_n \in \{\pm 1\}$ are drawn independently with $x_i \sim \rho$ for each $i \in [n]$.

The regularity condition we need is the following.

Definition 5 (Strong Distribution)**.** *Let ρ be a probability distribution on $[-1,1]$. Define $h_n^\rho : \{-n-1, -n+1, \cdots, n+1\} \to \mathbb{R}$ by*

$$h_n^\rho(t) = \frac{(n+1+t)(n+1-t)}{2(n+1)} \cdot \underset{p\sim\rho, x_{1\cdots n+1}\sim p}{\mathbb{P}}\left[\sum_{i\in[n+1]} x_i = t\right].$$

*We say ρ is (ξ, n)-*strong *if*

$$\sum_{t\in\{-n,-n+2,\cdots,n\}} \left|h_n^\rho(t-1) - h_n^\rho(t+1)\right| \le \xi.$$

*We say ρ is ξ-*strong *if ρ is (ξ, n)-strong for all n.*

We give some meaning to this technical definition in Section 2.4. Intuitively, it suffices for a distribution to have a "smooth" probability density function that is sufficiently "spread out." In particular, the uniform distribution on $[-1,1]$ is 1-strong.

Now we relate the definition of a strong distribution to the correlation quantity of interest:

Lemma 6. *Let ρ be a (ξ,n)-strong probability distribution on $[-1,1]$. Let $f:\{\pm 1\}^n \to [-1,1]$. Then*

$$\left| \mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[f(x)\sum_{i\in[n]}(x_i - z)\right]\right| \le \xi.$$

Furthermore, this is tight – that is, ρ is (ξ,n)-strong if and only if the above holds for all f.

Proof. Define a random variable S_n^ρ on $\{-n-1,-n+1,\cdots,n-1,n+1\}$ as follows. First sample $p\sim\rho$. Then sample $x_{1\cdots n+1}\sim p$ and let $S_n^\rho = \sum_{i=1}^{n+1} x_i$.

Firstly, by symmetry the following are equivalent ways of sampling random variables $x_1,\cdots,x_n,z$.

- Sample $p\sim\rho$. Then sample $x_{1\cdots n}\sim p$ and $z\sim p$.
- Sample $s\sim S_n^\rho$. Then sample $x_1,\cdots,x_n,z\in\{\pm 1\}$ uniformly at random conditioned on $z+\sum_{i\in[n]} x_i = s$.
- Sample $s\sim S_n^\rho$. Then sample $z\in\{\pm1\}$ with $\mathbb{E}[z]=\frac{s}{n+1}$. Then sample $x_1,\cdots,x_n\in\{\pm1\}$ uniformly at random conditioned on $\sum_{i\in[n]} x_i = s-z$.

Thus we can rewrite the expectation using S_n^ρ:

$$\begin{aligned}
\mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[f(x)\sum_{i\in[n]}(x_i - z)\right] &= \mathop{\mathbb{E}}_{s\sim S_n^\rho}\left[\mathop{\mathbb{E}}_{z\sim\frac{s}{n+1}}\left[\mathop{\mathbb{E}}_{x\in\{\pm1\}^n:\sum_{i\in[n]}x_i=s-z}\left[f(x)\sum_{i\in[n]}(x_i-z)\right]\right]\right]\\
&= \mathop{\mathbb{E}}_{s\sim S_n^\rho}\left[\mathop{\mathbb{E}}_{z\sim\frac{s}{n+1}}\left[\mathop{\mathbb{E}}_{x\in\{\pm1\}^n:\sum_{i\in[n]}x_i=s-z}[f(x)](s-z-nz)\right]\right]\\
&= \mathop{\mathbb{E}}_{s\sim S_n^\rho}\left[\mathop{\mathbb{E}}_{z\sim\frac{s}{n+1}}[g(s-z)(s-z-nz)]\right],
\end{aligned}$$

where $g:\{-n,\cdots,n\}\to[-1,1]$ given by

$$g(t) := \mathop{\mathbb{E}}_{x\in\{\pm1\}^n:\sum_{i\in[n]}x_i=t}[f(x)]$$

is the symmetrization of f. Now we expand the expectations as sums:

$$\begin{aligned}
\mathop{\mathbb{E}}_{s\sim S_n^\rho}\left[\mathop{\mathbb{E}}_{z\sim\frac{s}{n+1}}[g(s-z)(s-z-nz)]\right] &= \mathop{\mathbb{E}}_{s\sim S_n^\rho}\begin{bmatrix} \mathop{\mathbb{P}}_{z\sim\frac{s}{n+1}}[z=1] & \cdot & g(s-1)(s-1-n)\\ \mathop{\mathbb{P}}_{z\sim\frac{s}{n+1}}[z=-1] & \cdot & g(s+1)(s+1+n)\end{bmatrix}\\
&= \mathop{\mathbb{E}}_{s\sim S_n^\rho}\begin{bmatrix} \frac{n+1+s}{2(n+1)} & \cdot & g(s-1)(s-1-n)\\ \frac{n+1-s}{2(n+1)} & \cdot & g(s+1)(s+1+n)\end{bmatrix}\\
&= \mathop{\mathbb{E}}_{s\sim S_n^\rho}\left[\frac{(n+1+s)(n+1-s)}{2(n+1)}\cdot(g(s+1)-g(s-1))\right]\\
&=\frac{1}{2(n+1)}\sum_{s\in\{-n+1,-n+3,\cdots,n-1\}}\mathbb{P}\left[S_n^\rho=s\right]\cdot(n+1+s)(n+1-s)\cdot(g(s+1)-g(s-1))\\
&=\frac{1}{2(n+1)}\sum_{t\in\{-n+2,-n+4,\cdots,n\}}\mathbb{P}\left[S_n^\rho=t-1\right]\cdot(n+t)(n-t+2)\cdot g(t)\\
&\quad-\frac{1}{2(n+1)}\sum_{t\in\{-n,-n+2\cdots,n-2\}}\mathbb{P}\left[S_n^\rho=t+1\right]\cdot(n+t+2)(n-t)\cdot g(t)\\
&=\frac{1}{2(n+1)}\sum_{t\in\{-n,-n+2,\cdots,n\}}g(t)\cdot\begin{pmatrix}\mathbb{P}\left[S_n^\rho=t-1\right]\cdot(n+t)(n-t+2)\\ -\mathbb{P}\left[S_n^\rho=t+1\right]\cdot(n+t+2)(n-t)\end{pmatrix}\\
&=\sum_{t\in\{-n,-n+2,\cdots,n\}}g(t)\cdot(h(t-1)-h(t+1)),
\end{aligned}$$

where h is as in Definition 5. Now we can apply Hölder's inequality with the definitions of a strong distribution and g to conclude:

$$\begin{aligned}
\left|\mathop{\mathbb{E}}_{p\sim\rho,x_{1\cdots n}\sim p,z\sim p}\left[f(x)\sum_{i\in[n]}(x_i-z)\right]\right| &= \left|\mathop{\mathbb{E}}_{s\sim S_n^\rho}\left[\mathop{\mathbb{E}}_{z\sim\frac{s}{n+1}}[g(s-z)(s-z-nz)]\right]\right|\\
&=\left|\sum_{t\in\{-n,-n+2,\cdots,n\}}g(t)\cdot(h(t-1)-h(t+1))\right|\\
&\le\|g\|_\infty\sum_{t\in\{-n,-n+2,\cdots,n\}}|h(t-1)-h(t+1)|\\
&\le\xi.
\end{aligned}$$

Note that there exists a g that makes this inequality tight, namely

$$g_{\text{tight}}(t)=\operatorname{sign}(h(t-1)-h(t+1)).$$

Setting $f_{\text{tight}}(x)=g_{\text{tight}}\left(\sum_{i\in[n]}x_i\right)$ shows that the lemma is tight. □

Now we translate Lemma 6 into the form we will use:

Corollary 7. *Let ρ be a (ξ,n)-strong probability distribution on $[-1,1]$. Let $M:\{\pm1\}^n\to\mathbb{R}$ satisfy $|M(x)-\overline{x}|\le\alpha$ for all $x\in\{\pm1\}^n$. Then*

$$\mathop{\mathbb{E}}_{p\sim\rho,x_{1\cdots n}\sim p,z\sim p}\left[M(x)\sum_{i\in[n]}(x_i-z)\right]\ge\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right]-\alpha\xi.$$

Proof. Write $M(x) = \overline{x} - \alpha \cdot f(x)$ for some $f : \{\pm 1\}^n \to [-1,1]$. Now

$$\begin{aligned}
\mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[M(x)\sum_{i\in[n]}(x_i - z)\right] &= \mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[\overline{x}\sum_{i\in[n]}(x_i - z)\right] - \alpha \cdot \mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[f(x)\sum_{i\in[n]}(x_i - z)\right] \\
&\geq \mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[\overline{x}\sum_{i\in[n]}(x_i - z)\right] - \alpha \cdot \xi,
\end{aligned}$$

by Lemma 6. All that remains is a calculation:

$$\begin{aligned}
\mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[\overline{x}\sum_{i\in[n]}(x_i - z)\right] &= \mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p}[\overline{x}\cdot(\overline{x}-p)\cdot n] && (\text{since } \mathbb{E}[z] = p) \\
&= \mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p}[(\overline{x}-p)\cdot(\overline{x}-p)\cdot n] && (\text{since } \mathbb{E}[p\cdot(\overline{x}-p)] = 0) \\
&= \mathop{\mathbb{E}}_{p\sim\rho}\left[\mathop{\mathsf{Var}}_{x_{1\cdots n}\sim p}[\overline{x}]\cdot n\right] \\
&= \mathop{\mathbb{E}}_{p\sim\rho}\left[\mathop{\mathsf{Var}}_{x\sim p}[x]\right] \\
&= \mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right].
\end{aligned}$$

□

We now make an observation that will allow the construction of a high-power attack for symmetric M. Suppose $f : \{\pm 1\}^n \to [-1,1]$ can be written as $f(x) = f_*\left(\frac{1}{n}\sum_{i\in[n]} x_i\right)$ for some $f_* : [-1,1] \to [-1,1]$. Then, by symmetry, the conclusion of Corollary 7 can be altered to

$$\forall i \in [n] \quad \mathop{\mathbb{E}}_{p\sim\rho, x_1\cdots x_n\sim p, z\sim p}[f(x)\cdot(x_i - z)] \geq \frac{\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right] - \alpha\xi}{n}.$$

Formally, we have the following definition and Lemma.

Definition 8. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is* symmetric *if there exists a function $f_* : \mathbb{R} \to \mathbb{R}$ such that $f(x) = f_*\left(\frac{1}{n}\sum_{i\in[n]} x_i\right)$ for all $x \in \mathbb{R}^n$.*

Lemma 9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be symmetric and let $X_1, \cdots, X_n \in \mathbb{R}$ be independent and identically distributed. Then*

$$\mathop{\mathbb{E}}_{X}\left[f(X)(X_k - \mathbb{E}[X_k])\right] = \frac{1}{n}\mathop{\mathbb{E}}_{X}\left[f(X)\sum_{i\in[n]}(X_i - \mathbb{E}[X_i])\right]$$

for all $k \in [n]$.

Proof. By Definition 8,

$$\mathop{\mathbb{E}}_{X}\left[f(X)\sum_{i\in[n]}(X_i - \mathbb{E}[X_i])\right] = \sum_{i\in[n]}\mathop{\mathbb{E}}_{X}\left[f_*\left(\frac{1}{n}\sum_{k\in[n]} X_k\right)(X_i - \mathbb{E}[X_i])\right] \tag{1}$$

Since $X_1, \cdots, X_n$ are independent and identically distributed, the pair $(\sum_{k\in[n]} X_k, X_i)$ is identically distributed for all i. Thus $f_*(\frac{1}{n}\sum_{k\in[n]} X_k)(X_i - \mathbb{E}[X_i])$, being a function of $(\sum_{k\in[n]} X_k, X_i)$, is identically distributed for each i. Consequently, all the terms in (1) are the same, which implies the lemma. □

We can summarise our expectation bounds as follows.

Proposition 10. *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ, n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right] \geq \gamma + \alpha \cdot \xi$. Suppose the mechanism $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ is α-accurate. Let $x_1, \cdots x_n, z \sim \mathcal{P}_p$ and $q \sim \mathcal{M}(x_1, \cdots, x_n)$.*

1. *Then we have*

$$\forall j \in [d] \quad \mathop{\mathbb{E}}_{p,x_1,\cdots,x_n,z}\left[\sum_{i\in[n]}\left(\langle x_i^j, q^j\rangle - \langle z^j, q^j\rangle\right)\right] \geq \gamma.$$

   Moreover, this bound holds even when conditioned on all the randomness in columns other than j. That is, the bound holds when we condition on any value of $p^{-j}, \{x_i^{-j}\}_{i=1,\dots,n}, z^{-j}, q^{-j}$ and the randomness is only over the remaining variables.

2. *If, in addition, $\mathcal{M}$ is symmetric, then*

$$\forall j \in [d]\ \forall i \in [n] \quad \mathop{\mathbb{E}}_{p,x_1,\cdots,x_n,z}\left[\langle x_i^j, q^j\rangle - \langle z^j, q^j\rangle\right] \geq \frac{\gamma}{n}$$

   and hence

$$\forall i \in [n] \quad \mathop{\mathbb{E}}_{p,x_1,\cdots,x_n,z,\mathcal{M}}\left[\langle x_i, q\rangle - \langle z, q\rangle\right] \geq \frac{\gamma d}{n}.$$

Proof. We view z^{-j}, q^{-j}, x_i^{-j} as fixed and we average over the randomness of $\mathcal{M}$. Now the only randomness is the choice of p^j and $z^j, x_1^j \cdots x_n^j \sim p^j$. Since $\mathcal{M}$ does not see p^j or z^j, we can write $q^j = f(x^j)$ for some $f : \{\pm 1\}^n \to [-1,1]$. By the assumption that $\mathcal{M}$ is α-accurate, $|f(x) - \overline{x}| \leq \alpha$ for all $x \in \{\pm 1\}^n$. The result now follows from Corollary 7 and Lemma 9. □

2.3 Completeness Analysis

Now that we have shown that $\mathbb{E}\left[\sum_{i\in[n]}(\langle x_i, q\rangle - \langle z, q\rangle)\right]$ is large, we can turn this into a high probability statement.

Lemma 11. *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ, n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right] \geq \gamma + \alpha \cdot \xi$. Suppose the mechanism $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ is α-accurate. Assume $d > O(n^2 \log(1/\delta)/\gamma^2)$. Let $x_1, \cdots x_n, z \sim \mathcal{P}_p$ and $q \sim \mathcal{M}(x_1, \cdots, x_n)$. Then*

$$\mathop{\mathbb{P}}_{p,x_{1\cdots n},z,\mathcal{M}}\left[\sum_{i\in[n]}(\langle x_i, q\rangle - \langle z, q\rangle) < \frac{\gamma d}{2}\right] \leq \delta.$$

Moreover, if $\mathcal{M}$ is symmetric, then

$$\forall i \in [n] \quad \mathop{\mathbb{P}}_{p,x_{1\cdots n},z,\mathcal{M}}\left[\langle x_i, q\rangle - \langle z, q\rangle < \frac{\gamma d}{2n}\right] \leq \delta.$$

The formal proof of this Lemma is quite involved, but unenlightening. Thus we defer it to the appendix (page 34) and give a proof sketch here instead.

Proof Sketch. Write

$$\sum_{i\in[n]}(\langle x_i,q\rangle-\langle z,q\rangle)=\sum_{j\in[d]}q^j\cdot\sum_{i\in[n]}(x_i^j-z^j)=:\sum_{j\in[d]}A_j.$$

By Proposition 10, we have $\mathbb{E}\left[A_j\right]\geq\gamma$ for all $j\in[d]$. Suppose the A_j random variables were independent. Then we could apply Hoeffding's inequality. Using $|A_j|\leq 2n$, gives

$$\mathbb{P}\left[\sum_{j\in[d]}A_j-\mathbb{E}\left[A_j\right]<\frac{-1}{2}\gamma d\right]\leq\exp\left(-\frac{2(\gamma d/2)^2}{(4n)^2 d}\right)\leq\delta,$$

as required. The second half of the lemma is similar.

The A_j variables are not independent, but it turns out their sum concentrates nonetheless. The key observation is that $\mathbb{E}\left[A_j\right]\geq\gamma$ even if we condition on $A_1,\cdots,A_{j-1},A_{j+1},\cdots,A_d$. Namely

$$\mathbb{E}\left[A_j\mid A_1=a_1,\cdots,A_{j-1}=a_{j-1},A_{j+1}=a_{j+1},\cdots,A_d=a_d\right]\geq\gamma$$

for all $j\in[d]$ and $a\in\mathbb{R}^d$. □

Now we can finally prove completeness.

Proposition 12 (Completeness)**.** *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ,n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right]\geq\gamma+\alpha\cdot\xi$. Assume $d>O(n^2\log(1/\delta)/\gamma^2)$. Suppose the mechanism $\mathcal{M}:\{\pm1\}^{n\times d}\to[-1,1]^d$ is α-accurate. Let $x_1,\cdots,x_n,z\sim\mathcal{P}_p$ and $q=\mathcal{M}(x_1,\cdots,x_n)$. Then*

$$\mathop{\mathbb{P}}_{p,x_{1\cdots n},z,\mathcal{M}}[\exists i\in[n]\quad\mathcal{A}_{\delta,d}(x_i,q,z)=\text{IN}]\geq1-\delta.$$

Proof. By Lemma 11, $\sum_{i\in[n]}(\langle x_i,q\rangle-\langle z,q\rangle)\geq\frac{\gamma d}{2}>n\cdot\tau=n\cdot2\sqrt{d\log(1/\delta)}$ with high probability. Thus, with high probability, we have $\langle x_i,q\rangle-\langle z,q\rangle>\tau$ for at least one $i\in[n]$. □

We also state the high-power completeness we get from assuming that $\mathcal{M}$ is symmetric.

Proposition 13 (High-Power Completeness)**.** *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ,n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right]\geq\gamma+\alpha\cdot\xi$. Assume $d>O(n^2\log(1/\delta)/\gamma^2)$. Suppose the mechanism $\mathcal{M}:\{\pm1\}^{n\times d}\to[-1,1]^d$ is α-accurate and symmetric. Let $x_1,\cdots x_n,z\sim\mathcal{P}_p$ and $q\sim\mathcal{M}(x_1,\cdots,x_n)$. Then*

$$\forall i\in[n]\quad\mathop{\mathbb{P}}_{p,x_1,\cdots,x_n,z,\mathcal{M}}[\mathcal{A}_{\delta,d}(x_i,q,z)=\text{IN}]\geq1-\delta.$$

Proof. By Lemma 11, for all $i\in[n]$ we have $\langle x_i,q\rangle-\langle z,q\rangle\geq\frac{\gamma d}{2n}>\tau=2\sqrt{d\log(1/\delta)}$ with high probability. Thus, for all $i\in[n]$, we have $\langle x_i,q\rangle-\langle z,q\rangle>\tau$ with high probability. □

2.4 Interpreting Strong Distributions

The notion of strong distributions is critical in the completeness analysis of our attack – it ensures that the output of $\mathcal{M}$ correlates with its input. In this section we show that this condition is met by a large class of distributions and give some intuition for its meaning. First we restate Definition 5.

Definintion 5. *Let ρ be a probability distribution on $[-1,1]$. Define $h_n^\rho : \{-n-1, -n+1, \cdots, n+1\} \to \mathbb{R}$ by*

$$h_n^\rho(t) = \frac{(n+1+t)(n+1-t)}{2(n+1)} \cdot \underset{p\sim\rho, x_{1\cdots n+1}\sim p}{\mathbb{P}}\left[\sum_{i\in[n+1]} x_i = t\right].$$

*We say ρ is (ξ, n)-*strong *if*

$$\sum_{t\in\{-n,-n+2,\cdots,n\}} \left|h_n^\rho(t-1) - h_n^\rho(t+1)\right| \leq \xi.$$

*We say ρ is ξ-*strong *if ρ is (ξ, n)-strong for all n.*

First let us unpack this definition: The definition bounds the *total variation* of the function h_n^ρ. So we require h_n^ρ to be smooth. The function h_n^ρ is the product of two terms. The first term is large (at most $(n+1)/2$) in the middle of the range and smoothly decreases towards zero at the ends of the range. The second term can be viewed as the probability mass function of a discretization of the continuous distribution ρ: There are $n+2$ buckets $\{-n-1, -n+1, \cdots, n+1\}$. A sample $p \sim \rho$ is thrown into one of the $n+2$ buckets in a random fashion. The most likely bucket is the one closest to $(n+1)\cdot p$ and the probability of landing in a given bucket decays rapidly as we move away from the most likely bucket. Intuitively, being a strong distribution simply means that neighbouring buckets should contain a similar amount of probability mass.

To gain some intuition for the meaning of the definition, we consider some example distributions that do *not* satisfy the strong distribution assumption.

(i) Suppose ρ is a point mass on p^*. Then S_n^ρ is a (shifted and scaled) binomial distribution and h_n^ρ has high total variation. In this situation a simple mechanism $\mathcal{M}$ can prevent tracing: simply outputting $q = p^*$ will be accurate with high probability, but this allows the output of $\mathcal{M}$ to be (almost) independent from its input. Tracing is thus impossible, as there is almost no difference between the IN and OUT cases.

(ii) Example (i) can be generalised: Any distribution supported on a short interval is not strong.

(iii) Suppose ρ is supported on two points p^* and p^{**} that are far apart. Then S_n^ρ is a convex combination of shifted and scaled binomial distributions.

This corresponds to a mechanism $\mathcal{M}$ that knows p^* and p^{**} and returns one of the two if they are sufficiently accurate. Again, with high probability, the output of $\mathcal{M}$ is not sensitive to changes in the input. That means the output of $\mathcal{M}$ does not contain much information that is specific to its input. This makes tracing impossible.

(iv) Example (iii) can be generalised to any distribution supported on a small number of points. This can be generalised further to distributions supported on many short intervals.

The above examples demonstrate what a strong distribution avoids. Instead a strong distribution is "spread out" and "smooth."

The function h_n^ρ in Definintion 5 is somewhat unintuitive. We can give an alternative definition:

Lemma 14. *Let $U_1, \cdots, U_{n+1} \in [-1,1]$ be independent uniformly random variables and let $P \sim \rho$ be independent from $U_1, U_2, \cdots, U_{n+1}$. Let $U_{(1)} \geq U_{(2)} \geq \cdots \geq U_{(n)}$ denote the random variables in sorted order. Set $U_{(0)} = +1$ and $U_{(n+2)} = -1$. Then*

$$\underset{p\sim\rho, x_{1\cdots n+1}\sim p}{\mathbb{P}}\left[\sum_{i\in[n+1]} x_i = 2k-n-1\right] = \underset{U_0,\cdots,U_n,P}{\mathbb{P}}\left[U_{(k)} \geq P > U_{(k+1)}\right]$$

for all $k \in \{0, 1, \cdots, n+1\}$.

Thus the function h_n^ρ from Definition 5 can be defined as

$$h_n^\rho(t) = \frac{(n+1+t)(n+1-t)}{2(n+1)} \cdot \underset{U_0,\cdots,U_n,P}{\mathbb{P}}\left[U_{\left(\frac{t+n+1}{2}\right)} \geq P > U_{\left(\frac{t+n+3}{2}\right)}\right].$$

Intuitively, $U_{(0)} \leq U_{(1)} \leq \cdots \leq U_{(n+2)}$ partition the interval $[-1,1]$ into $n+2$ subintervals. Now h_n^ρ captures the amount of probability mass from ρ falling into each of these subintervals. However, the partitioning is itself random, so the probability mass at a particular point does not fall into a single subintervale. However, $U_{(k)} \approx \frac{n+2-2k}{n+2}$, so this random partitioning approximately partitions the interval evenly.

Proof of Lemma 14. Let $U_1, U_2, \cdots, U_{n+1}$ and P be sampled as in the lemma statement. Now define random variables $x_1, \cdots, x_{n+1} \in \{\pm 1\}$ by

$$x_i = 1 \iff U_i \geq P.$$

If we view P as fixed, then $\mathbb{P}[x_i = 1] = (P+1)/2$ and $\mathbb{E}[x_i] = P$ for each i. Moreover, the distribution of $x_1, \cdots, x_{n+1}$ is $n+1$ independent conditioned on P. We claim that, for any $k \in \{0, 1, \cdots, n+1\}$.,

$$\sum_{i\in[n+1]} x_i = 2k-n-1 \iff U_{(k)} \leq P \leq U_{(k+1)}.$$

The lemma follows from this claim, as we have shown a coupling between the two probability spaces under which the two events coincide.

To see the claim, note that $\sum_{i\in[n+1]} x_i = 2k-n-1$ if and only if k of the x_is are set to $+1$, which happens if and only if there are k choices of $i \in [n+1]$ with $U_i \geq P$. In turn this is equivalent to saying that the k^{th} largest U_i is greater than or equal to P, but the $(k+1)^{\text{th}}$ largest U_i is not — i.e. $U_{(k)} \geq P > U_{(k+1)}$. □

We can also characterise the limiting case (i.e. $n \to \infty$ rather than fixed n):

Proposition 15. *Let $\rho : [-1,1] \to \mathbb{R}$ be a continuously differentiable probability density function. Then ρ is a ξ-strong distribution if and only if*

$$\int_{-1}^{+1} \left|\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p)\right| \mathrm{d}p \leq \xi. \tag{2}$$

Proposition 15 shows that ρ being a strong probability density function is equivalent to a bound on the total variation of $(1-p^2)\rho(p)$. This function should be contrasted with h_n^ρ in Definition 5. Indeed, Proposition 15 is simply the result of taking $n \to \infty$ in Definition 5.

Proof of Proposition 15. Lemma 6 provides an exact characterisation of (ξ, n)-strong distributions. Namely, ρ is (ξ, n)-strong if and only if

$$\left| \underset{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}{\mathbb{E}} \left[f(x) \sum_{i\in[n]} (x_i - z) \right] \right| \le \xi \tag{3}$$

for all $f : \{\pm 1\}^n \to [-1,1]$. To show that ρ is ξ-strong we must show that (3) holds for all n and all f.

Fix n and $f : \{\pm 1\}^n \to [-1,1]$. Define $g : [-1,1] \to [-1,1]$ by

$$g(p) = \underset{x_{1\cdots n}\sim p}{\mathbb{E}} [f(x)].$$

By Lemma 16, for any $p \in [-1,1]$,

$$\underset{x_{1\cdots n}\sim p}{\mathbb{E}} \left[f(x) \cdot \sum_{i\in[n]} (x_i - p) \right] = g'(p) \cdot (1-p^2).$$

Thus

$$\begin{aligned} \underset{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}{\mathbb{E}} \left[f(x) \sum_{i\in[n]} (x_i - z) \right] &= \underset{p\sim\rho}{\mathbb{E}} \left[g'(p)(1-p^2) \right] \\ &= \int_{-1}^{+1} g'(p)(1-p^2)\rho(p)\mathrm{d}p. \end{aligned}$$

Now we apply integration by parts — that is, we integrate both sides of an application of the differentiation product rule:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}p} g(p)(1-p^2)\rho(p) =& g'(p)(1-p^2)\rho(p) + g(p)\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p). \\ \int_{-1}^{+1} \frac{\mathrm{d}}{\mathrm{d}p} g(p)(1-p^2)\rho(p)\mathrm{d}p =& \int_{-1}^{+1} g'(p)(1-p^2)\rho(p)\mathrm{d}p + \int_{-1}^{+1} g(p)\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p)\mathrm{d}p. \\ \int_{-1}^{+1} g'(p)(1-p^2)\rho(p)\mathrm{d}p =& \Big(g(1)(1-1^2)\rho(1) - g(-1)(1-(-1)^2)\rho(-1) \Big) \\ & - \int_{-1}^{+1} g(p)\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p)\mathrm{d}p. \end{aligned}$$

Finally, we can apply Hölder's inequality:

$$\begin{aligned}\left|\mathop{\mathbb{E}}_{p\sim\rho,x_{1\cdots n}\sim p,z\sim p}\left[f(x)\sum_{i\in[n]}(x_i-z)\right]\right| &= \left|\int_{-1}^{+1} g'(p)(1-p^2)\rho(p)\mathrm{d}p\right| \\ &= \left|\int_{-1}^{+1} g(p)\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p)\mathrm{d}p\right| \\ &\leq \|g\|_\infty \cdot \int_{-1}^{+1}\left|\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p)\right|\mathrm{d}p \\ &\leq \int_{-1}^{+1}\left|\frac{\mathrm{d}}{\mathrm{d}p}(1-p^2)\rho(p)\right|\mathrm{d}p.\end{aligned}$$

This proves one side of the equivalence. The other side of the equivalence follows from the tightness of Hölder's inequality and the fact that, by choosing n large enough, we can make $g:[-1,1]\to[-1,1]$ arbitrarily close to the function that makes the inequality tight. □

Now we give a technical lemma needed in the above proof. This result is similar to [SU14, Lemma 2.11]. (It can be viewed as a rescaling of said lemma.)

Lemma 16. *Let* $f:\{\pm1\}^n\to\mathbb{R}$. *Define* $g:[-1,1]\to\mathbb{R}$ *by*

$$g(p) = \mathop{\mathbb{E}}_{x_{1\cdots n}\sim p}[f(x)].$$

Then

$$\mathop{\mathbb{E}}_{x_{1\cdots n}\sim p}\left[f(x)\cdot\sum_{i\in[n]}(x_i-p)\right] = g'(p)\cdot(1-p^2).$$

Proof. Since $x^2=1$ for $x\in\{\pm1\}$, we have the identity

$$\frac{\mathrm{d}}{\mathrm{d}p}\frac{1+xp}{2} = \frac{x}{2} = \frac{1+xp}{2}\frac{x-p}{1-p^2}$$

for all $x\in\{\pm1\}$ and $p\in(-1,1)$. By the product rule, we have

$$\frac{\mathrm{d}}{\mathrm{d}p}\prod_{i\in[n]}\frac{1+x_ip}{2} = \sum_{i\in[n]}\left(\frac{\mathrm{d}}{\mathrm{d}p}\frac{1+x_ip}{2}\right)\prod_{k\in[n]\setminus\{i\}}\frac{1+x_kp}{2} = \sum_{i\in[n]}\frac{x_i-p}{1-p^2}\prod_{k\in[n]}\frac{1+x_kp}{2}$$

for all $x\in\{\pm1\}^n$ and $p\in(-1,1)$. Sampling $x\sim p$ samples each $x\in\{\pm1\}$ with probability $\frac{1+xp}{2}$. Thus sampling $x_{1\cdots n}\sim p$, samples each $x\in\{\pm1\}^n$ with probability $\prod_{i\in[n]}\frac{1+x_ip}{2}$.

Now we can write

$$g(p) = \mathop{\mathbb{E}}_{x_{1\cdots n}\sim p}[f(x)] = \sum_{x\in\{\pm1\}^n} f(x)\prod_{i\in[n]}\frac{1+x_ip}{2}.$$

Using the above identities gives

$$\begin{aligned} g'(p) &= \sum_{x \in \{\pm 1\}^n} f(x) \frac{\mathrm{d}}{\mathrm{d}p} \prod_{i \in [n]} \frac{1+x_i p}{2} \\ &= \sum_{x \in \{\pm 1\}^n} f(x) \sum_{i \in [n]} \frac{x_i - p}{1-p^2} \prod_{k \in [n]} \frac{1+x_k p}{2} \\ &= \mathop{\mathbb{E}}_{x_{1\cdots n} \sim p} \left[f(x) \sum_{i \in [n]} \frac{x_i - p}{1-p^2} \right] \end{aligned}$$

Rearranging gives the result. □

Using the differentiation product rule and the triangle inequality, we can show that

$$\begin{aligned} \int_{-1}^{+1} \left| \frac{\mathrm{d}}{\mathrm{d}p} (1-p^2)\rho(p) \right| \mathrm{d}p &= \int_{-1}^{+1} \left| (1-p^2)\rho'(p) - 2p\rho(p) \right| \mathrm{d}p \\ &\le \int_{-1}^{+1} (1-p^2) \left| \rho'(p) \right| \mathrm{d}p + \int_{-1}^{+1} \left| 2p\rho(p) \right| \mathrm{d}p \\ &\le \int_{-1}^{+1} \left| \rho'(p) \right| \mathrm{d}p + 2. \end{aligned}$$

Thus, rather than bounding the total variation of $(1-p^2)\rho(p)$, it suffices to bound the total variation of ρ.

A bound on the total variation of the probability density function is a very natural "smoothness" condition. In particular, the uniform distribution, whose probability density function is the constant $\frac{1}{2}$, has zero total variation. Thus Proposition 15 justifies our assertion that strong distributions correspond to a smoothness condition.

Using Proposition 15 we can give some examples of strong distributions:

- The uniform distribution on $[-1,1]$ is 1-strong.
- The uniform distribution on $[a,b]$ is ξ-strong for

$$\xi = \frac{2 - a^2 - b^2 + \int_a^b |2x| \mathrm{d}x}{b-a} \le \frac{2}{b-a} + 2.$$

- The (scaled) Beta distribution, with $\rho(p) \propto (1+p)^{u-1}(1-p)^{v-1}$ (where $u > 0$ and $v > 0$ and the support is $[-1,1]$), is $(4uv/(u+v))$-strong.

3 Tracing from Fewer Statistics

In the previous section we focused on tracing from very weak assumptions—weakly accurate answers and only a single reference sample. The price of these weak assumptions is that we (provably) need $d = \Omega(n^2)$ and can only trace a single individual. In this section we show that if the mechanism gives more accurate answers, then we can trace with smaller d, and can trace many individuals in the dataset. In exchange, we require a larger reference sample. More precisely, we show that if the mechanism is α-accurate (for some $\alpha \geq n^{-1/2}$), and we are given roughly $1/\alpha^2$ independent reference samples from the distribution, then we can trace when the dataset has dimension only $O(\alpha^2 n^2)$, and we can successfully trace $\Omega(1/\alpha^2)$ individuals in the dataset. We summarise our results in the following informal theorem, which effectively generalises Theorem 1 from the introduction.

Theorem 17 (Informal)**.** *For every $\delta > 0$, $n \in \mathbb{N}$, $\alpha \geq 1/n^{1/2}$, $d \geq O(\alpha^2 n^2 \log(1/\delta))$, $m \geq O(\log(n)/\alpha^2)$, and $t \leq \Omega(1/\alpha^2)$, there exists an attacker $\mathcal{A}^* : \{\pm 1\}^d \times [\pm 1]^d \times (\{\pm 1\}^d)^{m+1} \to \{\mathrm{IN}, \mathrm{OUT}\}$ the following holds.*

Let $\mathcal{D}$ be a product distribution on $[-1,1]^d$ such that each marginal satisfies a technical smoothness condition (Definition 5). Let $\mathcal{M} : \{\pm 1\}^{n \times d} \to [-1,1]^d$ be α-accurate. Let $p \sim \mathcal{D}$ and $x_1, \cdots, x_n, y, z_0, z_1, \ldots, z_m \sim \mathcal{P}_p$. Let $q \sim \mathcal{M}(x_1, \cdots, x_n)$. Then

$$\mathbb{P}[\mathcal{A}^*(y, q, (z_0, z_1, \ldots, z_m)) = \mathrm{IN}] \leq \delta, \textit{and}$$

$$\mathbb{P}[|\{i \in [n] \mid \mathcal{A}^*(x_i, q, (z_0, z_1, \ldots, z_m)) = \mathrm{IN}\}| \geq t] \geq 1 - \delta.$$

The modified attack is described below. In the attack, y represents the targeted individual, q is a vector of the mechanism's answers, and $z_0, z_1, \ldots, z_m$ represent $m+1$ independent reference samples from the distribution. The first reference sample z_0 is used exactly as before as an unbiased estimate of p. The remaining m samples $z_1, \ldots, z_m$ will be averaged to form an independent unbiased estimate of p with much lower variance. We will set $m \approx 1/\alpha^2$ so that this estimate is α-accurate.

```
1. Input: $y, z_0, z_1, \ldots, z_m \in \{\pm 1\}^d$, and $q \in [\pm 1]^d$.
2. Let $z = z_0$ and $w = (1/m) \sum_{i=1}^m z_i$.
3. Let $\eta := 2\alpha$ and let $\lfloor q - w \rceil_\eta \in [-\eta, \eta]^d$ be the entrywise truncation of $q - w$, to $[-\eta, \eta]$. (We
   believe that this truncation is unnecessary, but it is needed for our analysis.)
4. Compute
   $$\langle y - z, \lfloor q - w \rceil_\eta \rangle = \sum_{j \in [d]} (y^j - z^j) \cdot \left\lfloor q^j - w^j \right\rceil_\eta .$$
5. If $\langle y - z, \lfloor q - w \rceil_\eta \rangle > \tau := 4\alpha \sqrt{d \log(1/\delta)}$, output IN; otherwise output OUT.
```

Figure 2: Attack with a Large Reference Sample $\mathcal{A}^*_{\delta,\alpha,d,m}(y, q, \vec{z})$

3.1 Soundness

Proposition 18 (Soundness)**.** *Fix any $q, z_1, \ldots, z_m, p \in [-1,1]^d$. Suppose $y, z_0 \sim \mathcal{P}_p$ are independent from each other and from $q, z_1, \ldots, z_m$. Then*

$$\mathbb{P}\left[\mathcal{A}^*_{\delta,\alpha,d,m}(y, q, \vec{z}) = \mathrm{IN}\right] \leq \delta.$$

Proof. Since y and z_0 are identically distributed, and $q, z_1, \ldots, z_m$ are fixed

$$\mathbb{E}\left[\langle y - z, \lfloor q - w \rceil_\eta \rangle\right] = 0$$

(recall $z = z_0$ and $w = (1/m)\sum_{i=1}^m z_i$). Since y and z_0 are independent samples from a product distribution, we have that $\langle y - z, \lfloor q - w \rceil_\eta \rangle = \sum_{i \in [d]} (y^j - z^j) \cdot \lfloor q - w \rceil_\eta^j$ is the sum of $2d$ independent random variables, each of which is bounded by $\eta = 2\alpha$. Thus, by Hoeffding's inequality,

$$\mathbb{P}\left[\langle y - z, \lfloor q - w \rceil_\eta \rangle > \tau\right] \leq e^{-\tau^2/16d\alpha^2} \leq \delta.$$

This completes the proof. □

3.2 Correlation Analysis

We have the following proposition, analogous to Proposition 10 in Section 2.2.

Lemma 19. *Let $\mathcal{M} : \{\pm 1\}^{n \times d} \to [-1,1]^d$ be α-accurate, let $\eta = 2\alpha$, and let the distribution $\mathcal{D}$ be a product distribution where every marginal ρ is (ξ, n)-strong and satisfies $\mathop{\mathbb{E}}_{p \sim \rho}\left[1 - p^2\right] \geq \gamma + \alpha\xi$. Consider the following experiment. Let $p \sim \mathcal{D}$, let $x_1, \ldots, x_n, z_0, z_1, \ldots, z_m \sim \mathcal{P}_p$, and $q \sim \mathcal{M}(x_1, \ldots, x_n)$. Then for every $j \in [d]$,*

$$\mathbb{E}\left[\sum_{i \in [n]} (x_i^j - z^j) \lfloor q - w \rceil_\eta^j\right] \geq \gamma - 4n \cdot e^{-\alpha^2 m/2},$$

where $z = z_0$ and $w = (1/m)\sum_{i=1}^m w_i$.

Moreover, this statement holds even when we condition on everything pertaining to columns other than j. That is, the bound on the expectation holds when we condition on any value of $p^{-j}, \{x_i^{-j}\}_{i=1,\ldots,n}, \{z_i^{-j}\}_{j=0,1,\ldots,m}$, and q^{-j} and the randomness is taken only over the remaining variables.

Proof. Since $\mathcal{M}$ is α-accurate and the distribution is (ξ, n)-strong, by Proposition 10

$$\mathbb{E}\left[\sum_{i \in [n]} (x_i^j - z^j) \cdot (q^j - w^j)\right] \geq \gamma.$$

So it remains to show that

$$\mathbb{E}\left[\sum_{i \in [n]} (x_i^j - z^j)(q^j - w^j - \lfloor q - w \rceil_\eta^j)\right] \leq 4ne^{-\alpha^2 m/2}.$$

Since $\left|\sum_{i \in [n]} (x_i^j - z^j) \cdot (q^j - w^j - \left\lfloor q^j - w^j \right\rceil_\eta)\right| \leq 4n$ and $\sum_{i \in [n]} (x_i^j - z^j)(q^j - w^j - \left\lfloor q^j - w^j \right\rceil_\eta) = 0$ when $|q^j - w^j| \leq \eta$, it suffices to show that $\mathbb{P}\left[|q^j - w^j| > \eta\right] \leq e^{-\alpha^2 m/2}$. By accuracy, we have $|q^j - p^j| \leq \alpha$, and by a Chernoff bound, we have $\mathbb{P}\left[|p^j - w^j| > \alpha\right] \leq e^{-\alpha^2 m/2}$. This completes the proof. □

Proposition 20. *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ, n)-strong and satisfies $\mathop{\mathbb{E}}_{p \sim \rho}\left[1 - p^2\right] \geq \gamma + \alpha\xi$. Suppose $\mathcal{M} : \{\pm 1\}^{n \times d} \to [-1,1]^d$ is α-accurate. Let $d > O(\alpha^2 n^2 \log(1/\delta)/\gamma^2)$ and $m \geq 2\log(24n/\gamma)/\alpha^2$. Let $x_1, \ldots, x_n, z_0, z_1, \ldots, z_m \sim \mathcal{P}_p$. Let $q \sim \mathcal{M}(x_1, \ldots, x_n)$ Then*

$$\mathbb{P}\left[\sum_{i \in [n]} \left(\langle x_i - z, \lfloor q - w \rceil_\eta \rangle\right) < \frac{\gamma d}{2}\right] \leq \delta$$

(recall $z = z_0$, $w = (1/m)\sum_{i=1}^m z_i$, and $\eta = 2\alpha$).

The proof of Proposition 20 is analogous to that of Lemma 11 and is presented in Section A.2.

Proposition 20 establishes a lower bound on the sum of the expected scores. Next we will upper bound the 2-norm of the expected scores. Upper bounding the 2-norm will establish that the scores are "spread out," so there must be many (roughly $1/\alpha^2$) expected scores that are large (larger than the threshold τ).

Our analysis relies on the following technical lemma.

Lemma 21. *Let $X_1, \cdots, X_n \in \mathbb{R}$ be independent random variables such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\left[X_i^2\right] \le 1$ for every $i \in [n]$. Let $Y \in \mathbb{R}$ be another (not necessarily independent) random variable. Then*

$$\sum_{i\in[n]} \mathbb{E}[X_i Y]^2 \le \mathbb{E}\left[Y^2\right].$$

Proof. For $i \in [n]$, let $c_i = \mathbb{E}[X_i Y]$. Define $h : \mathbb{R}^n \to \mathbb{R}$ by $h(x) = \sum_{i\in[n]} c_i x_i$. Then

$$\mathbb{E}\left[h(X)^2\right] = \sum_{i,j\in[n]} c_i c_j \mathbb{E}\left[X_i X_j\right] \le \sum_{i\in[n]} c_i^2$$

and

$$\mathbb{E}[h(X)Y] = \sum_{i\in[n]} c_i \mathbb{E}[X_i Y] = \sum_{i\in[n]} c_i^2.$$

Thus

$$0 \le \mathbb{E}\left[(h(X)-Y)^2\right] = \mathbb{E}\left[h(X)^2\right] - 2\mathbb{E}[h(X)Y] + \mathbb{E}\left[Y^2\right] \le \sum_{i\in[n]} c_i^2 - 2\sum_{i\in[n]} c_i^2 + \mathbb{E}\left[Y^2\right].$$

Rearranging gives

$$\sum_{i\in[n]} c_i^2 \le \mathbb{E}\left[Y^2\right],$$

as required. □

Lemma 22. *Fix $p \in [-1,1]^d$ and let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ be any mechanism. Fix any w and let $x_1, \cdots, x_n, z_0 \sim \mathcal{P}_p$ and $q \sim \mathcal{M}(x_1, \cdots, x_n)$. Then for every $j \in [d]$,*

$$\sqrt{\sum_{i\in[n]} \mathbb{E}\left[\langle x_i^j - z^j, \left\lfloor q^j - w^j \right\rceil_\eta \rangle\right]^2} \le \eta\sqrt{2}$$

(recall $z = z_0$). Moreover, this statement holds even when we condition on everything pertaining to columns other than j. That is, the bound holds when we condition the expectations on any value of $\{x_i^{-j}\}_{i=1,\ldots,n}, z_0^{-j}$, and q^{-j} and the randomness is taken only over the remaining variables.

Proof. We apply Lemma 21 with $X_i = x_i^j - z^j$ and $Y = \left\lfloor q^j - w^j \right\rceil_\eta$. □

Once again, we would like to apply a concentration result to turn our bound on the sum of the squares of the expected scores into a high confidence bound on the sum of the squares of the scores themselves. Once again, this issue is complicated by a lack of independence. Nonetheless, we prove a suitable concentration bound for the sum of the squares of the scores in Proposition 46. Using this concentration bound we can prove the following.

Proposition 23. *Fix $p \in [-1,1]^d$ and let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ be any mechaniam. Assume $d \geq 64(n + \sqrt{\log(1/\delta)})$. Let $x_1, \cdots, x_n, z_0, z_1, \cdots, z_m \sim \mathcal{P}_p$, and let $q \sim \mathcal{M}(x_1, \cdots, x_n)$. Then*

$$\mathbb{P}\left[\sqrt{\sum_{i\in[n]} \langle x_i - z, \lfloor q - w \rceil_\eta \rangle^2} \leq 2\eta d\right] \geq 1 - \delta$$

(recall $z_0 = z$ and $w = (1/m)\sum_{i=1}^n z_i$).

Proof. By applying the triangle inequality to Lemma 22, we have

$$\sqrt{\sum_{i\in[n]} \mathbb{E}\left[\langle x_i - z, \lfloor q - w \rceil_\eta \rangle\right]^2} \leq d\eta\sqrt{2}.$$

By Theorem 46, for any $\lambda > 0$,

$$\mathbb{P}\left[\sqrt{\sum_{i\in[n]} \langle x_i - z, \lfloor q - w \rceil_\eta \rangle^2} > \lambda + d\eta\sqrt{2}\right] \leq \exp\left(\frac{nd}{2} - \frac{\lambda^2}{16\eta^2}\right).$$

The theorem follows by setting $\lambda = 4\eta\sqrt{\frac{nd}{2} + \log(1/\delta)} \leq \frac{\eta d}{2}$. □

Combining Proposition 20 with Proposition 23, we can show that, with high probability, the attack says IN for many target individuals x_i. To do so, we need the following elementary lemma.

Lemma 24. *Let $\sigma \in \mathbb{R}^n$ satisfy $\sum_{i\in[n]} \sigma_i \geq A$ and $\sum_{i\in[n]} \sigma_i^2 \leq B^2$. Then*

$$\left|\left\{i \in [n] : \sigma_i > \frac{A}{2n}\right\}\right| \geq \left(\frac{A}{2B}\right)^2.$$

Proof. Let $\tau = A/2n$ and $S = \{i \in [n] : \sigma_i > \tau\}$. Let $\sigma_S \in \mathbb{R}^{|S|}$ denote the restriction of σ onto the coordinates indexed by S. Then

$$\begin{aligned} A \leq \sum_{i\in[n]} \sigma_i = \sum_{i\in[n]\setminus S} \sigma_i + \sum_{i\in S} \sigma_i \\ &\leq (n - |S|)\tau + \|\sigma_S\|_1 \\ &\leq n\tau + \sqrt{|S|}\cdot\|\sigma_S\|_2 \\ &\leq n\tau + \sqrt{|S|}\cdot\|\sigma\|_2 \\ &\leq n\tau + \sqrt{|S|}\cdot B. \end{aligned}$$

Rearranging gives

$$|S| \geq \left(\frac{A - n\tau}{B}\right)^2 = \left(\frac{A}{2B}\right)^2,$$

as required. □

Proposition 25 (Completeness with a Large Reference Sample). *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ, n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right] \geq \gamma + \alpha\xi$. Suppose$\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ is α-accurate. Let $d > O(\alpha^2 n^2 \log(1/\delta)/\gamma^2)$ and $m \geq 2\log(24n/\gamma)/\alpha^2$. Let $x_1, \dots, x_n, z_0, z_1, \dots, z_n \sim \mathcal{P}_p$. Let $q \sim \mathcal{M}(x_1, \dots, x_n)$. Then*

$$\mathbb{P}\left[\left|\left\{i \in [n] : \mathcal{A}^*_{\delta,\alpha,d,m}(x_i, q, \vec{z}) = \mathrm{IN}\right\}\right| \geq \frac{\gamma^2}{256\alpha^2}\right] \geq 1 - 2\delta.$$

Proof. By Proposition 20, with probability at least $1-\delta$,

$$\sum_{i\in[n]} \left(\langle x_i - z, \lfloor q - w \rceil_\eta \rangle\right) \geq \frac{\gamma d}{2} =: A.$$

By Proposition 23, with probability at least $1-\delta$,

$$\sqrt{\sum_{i\in[n]} \langle x_i - z, \lfloor q - w \rceil_\eta \rangle^2} \leq 2\eta d =: B.$$

By a union bound, both of these events occur with probability at least $1-2\delta$. Assuming they both occur, Lemma 24 implies

$$\left|\left\{i \in [n] : \langle x_i - z, \lfloor q - w \rceil_\eta \rangle \geq \frac{A}{2n}\right\}\right| \geq \left(\frac{A}{2B}\right)^2 = \left(\frac{\gamma}{16\alpha}\right)^2.$$

We have $A/2n = \gamma d/4n \geq \tau = 4\alpha\sqrt{d\log(1/\delta)}$, which implies the result. □

4 Extensions

4.1 Robustness: Mechanisms with ℓ_1-Bounded Error

We have taken $\mathcal{M}$ being accurate to mean $\left\|\mathbb{E}[q] - p\right\|_\infty \leq \alpha$ for all p, where $q = \mathcal{M}(x)$ and the expectation is taken over the randomness of $\mathcal{M}$ and x. This condition is quite strong. Ideally, we would only need to assume, say, $\left\|\mathbb{E}[q] - p\right\|_1 \leq \alpha d$ – a very weak average-case error guarantee.

To achieve this, we must alter the definition of a strong distribution:

Definition 26 (Robustly Strong Distribution). *A probability distribution ρ on $[-1,1]$ is (η, γ)-*robustly strong *if*

$$\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2) + \frac{1}{\eta}|g(p) - p|\right] \geq \gamma$$

for any polynomial $g : [-1,1] \to [-1,1]$.

It can be verified that the uniform distribution is $(1/2, 1/3)$-robustly strong.

Soundness holds as before, but Completeness can be strengthened to the following.

Proposition 27 (Robust Completeness). *Suppose the distribution $\mathcal{D}$ is a product distribution on $[-1,1]^d$ in which each marginal is (η, γ)-robustly strong. Assume $d > O(n^2 \log(1/\delta)/\gamma^2)$. Let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$. Let $p \sim \mathcal{D}$, $x_1, \cdots x_n, z \sim \mathcal{P}_p$, and $q = \mathcal{M}(x_1, \cdots, x_n)$. Then*

$$\mathbb{P}\left[\|q - p\|_1 > \alpha d \quad \vee \quad \exists i \in [n] \;\; \mathcal{A}_{\delta,d}(x_i, q, z) = \mathrm{IN}\right] \geq 1 - \delta.$$

Proposition 27 follows from the following analogs of Lemmas **??** and 11.

Lemma 28. *Let $f : \{\pm 1\}^n \to [-1,1]$. Let ρ be a (η,γ)-robustly strong probability distribution. Then*

$$\mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[f(x) \sum_{i\in[n]} (x_i - z) + \frac{1}{\eta} |f(x) - p| \right] \geq \gamma.$$

Proof. Define $g : [-1,1] \to [-1,1]$ by $g(p) = \mathop{\mathbb{E}}_{x_{1\cdots n}\sim p} [f(x)]$. By Lemma 16 ,

$$\mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p}\left[f(x) \sum_{i\in[n]} (x_i - z) \right] = \mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p) \cdot (1-p^2) \right].$$

By Convexity,

$$\mathop{\mathbb{E}}_{p\sim\rho, x_{1\cdots n}\sim p, z\sim p} [|f(x) - p|] \geq \mathop{\mathbb{E}}_{p\sim\rho} [|g(p) - p|].$$

The lemma now follows from Definition 26. □

Lemma 29. *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal is (η,γ)-robustly strong. Let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$. Let $p \sim \mathcal{D}$, $x_1, \cdots x_n, z \sim \mathcal{P}_p$, and $q = \mathcal{M}(x_1, \cdots, x_n)$. Assume $d > O(n^2 \log(1/\delta)/\gamma^2)$. Then*

$$\mathop{\mathbb{P}}_{p, x_{1\cdots n}, z, \mathcal{M}}\left[\sum_{i\in[n]} \langle x_i, q\rangle - \langle z, q\rangle + \frac{1}{\eta} \|q - p\|_1 \geq \frac{1}{2}\gamma d \right] \geq 1 - \delta.$$

The proof is deferred to the appendix.

4.2 Generalizations to Real-Valued Data

The results of the previous sections generalize nearly directly to Gaussian data with a fixed variance. Specifically, suppose that the data $X \in \mathbb{R}^{n\times d}$ is drawn independently with $x_i^j \sim N(\mu_j, \sigma_j^2)$, where μ_j, σ_j are themselves random variables distributed over $[-1,1]$ and $[0, \sigma_{\max}]$ respectively according to a product distribution.

The attack is modified slightly in Figure 3.

```
$\mathcal{A}'_{\delta,d,\sigma_{\max}}(y,q,z)$

1. Input: $y, z \in \mathbb{R}^d$ and $q \in [-1,1]^d$.

2. Compute $\langle y, q\rangle = \sum_{j\in[d]} y^j \cdot q^j$ and $\langle z, q\rangle = \sum_{j\in[d]} z^j \cdot q^j$.

3. If $\langle y, q\rangle - \langle z, q\rangle > \tau' := 2\sigma_{\max}\sqrt{d \ln(1/\delta)}$, output IN; otherwise output OUT.
```

Figure 3: Our Privacy Attack for Real-Valued Data

4.2.1 Soundness

Verifying soundness of our attack is again straightforward.

Proposition 30 (Soundness)**.** *Let $q, \mu \in [-1,1]^d$ and $\sigma \in [0, \sigma_{\max}]^d$. Suppose $y, z \sim N(\mu, \mathrm{diag}(\sigma)^2)$ are independent from each other and from q.*[3] *Then*

$$\mathbb{P}\left[\mathcal{A}'_{\delta,d,\sigma_{\max}}(y,q,z) = \mathrm{IN}\right] \leq \delta.$$

Proof. We have that $y - z \sim N(0, 2 \cdot \mathrm{diag}(\sigma)^2)$. Thus $\langle y, q\rangle - \langle z, q\rangle \sim N(0, 2\sum_{j\in[d]} \sigma_j^2 q_j^2)$. Since $2\sum_{j\in[d]} \sigma_j^2 q_j^2 \leq 2d\sigma_{\max}^2$, we have

$$\mathbb{P}[\langle y,q\rangle - \langle z,q\rangle > \tau'] \leq \frac{1}{2}\exp\left(\frac{-\tau'^2}{2 \cdot 2d\sigma_{\max}^2}\right) \leq \delta,$$

as required. □

4.2.2 Proving Correlation

Now we can prove an analog to Lemma 16:

Lemma 31. *Let $f : \mathbb{R}^n \to [-1,1]$ be a (measurable) function. Define $g : \mathbb{R} \times (0,\infty) \to \mathbb{R}$ by*

$$g(\mu,\sigma) = \mathop{\mathbb{E}}_{x \sim N(\mu\vec{1}_n, \sigma^2 I_n)}[f(x)].$$

Then

$$\mathop{\mathbb{E}}_{x \sim N(\mu\vec{1}_n, \sigma^2 I_n)}\left[f(x)\sum_{i\in[n]}(x_i - \mu)\right] = \sigma^2\left(\frac{\partial}{\partial\mu} g(\mu,\sigma)\right).$$

In Lemma 16, we get $\mathop{\mathbb{E}}_{x_{1\cdots n}\sim p}\left[f(x)\cdot\sum_{i\in[n]}(x_i - p)\right] = g'(p)\cdot(1-p^2)$. The variance of $x \sim p$ is $1 - p^2$. Thus there is a very close connection between Lemmas 16 and 31.

Proof. We have

$$\begin{aligned}
\frac{\partial}{\partial\mu} g(\mu,\sigma) &= \frac{\partial}{\partial\mu} \mathop{\mathbb{E}}_{x \sim N(\mu\vec{1}_n, \sigma^2 I_n)}[f(x)] \\
&= \frac{\partial}{\partial\mu}\int_{\mathbb{R}^n} f(x)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i\in[n]}(x_i-\mu)^2}\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \left(\frac{\partial}{\partial\mu} e^{\frac{-1}{2\sigma^2}\sum_{i\in[n]}(x_i-\mu)^2}\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x)\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\frac{-1}{2\sigma^2}\sum_{i\in[n]}(x_i-\mu)^2}\frac{1}{\sigma^2}\sum_{i\in[n]}(x_i-\mu)\mathrm{d}x \\
&= \mathop{\mathbb{E}}_{x \sim N(\mu\vec{1}_n, \sigma^2 I_n)}\left[f(x)\frac{1}{\sigma^2}\sum_{i\in[n]}(x_i-\mu)\right].
\end{aligned}$$

[3] $x \sim N(\mu, \mathrm{diag}(\sigma)^2)$ denotes that each x_j is drawn independently from a Gaussian distribution with mean μ_j and variance σ_j^2.

Rearranging gives the result. □

The relevant notion of smoothness is now the following.

Definition 32. *A distribution ρ on pairs $(\mu,\sigma) \in [-1,1] \times [0,\sigma_{\max}]$ is* (α,γ)-strong for Gaussians *if for all continuously differentiable functions $g : [-1,1] \times [0,\sigma_{\max}] \to [-1,1]$ such that $|g(\mu,\sigma) - \mu| \le \alpha$ for all $\mu \in [-1,1]$ and $\sigma \in [0,\sigma_{\max}]$, we have*

$$\underset{(\mu,\sigma)\sim\rho}{\mathbb{E}}\left[\sigma^2 \frac{\partial}{\partial\mu} g(\mu,\sigma)\right] \ge \gamma.$$

When σ is constant and μ is uniform on an interval $[a,b]$, then this definition is satisfied with $\gamma = 1 - \frac{2\alpha}{b-a}$. Note that, unlike the boolean case, the mean and variance of a Normal are not necessarily related. This means we can consider the simpler case where the variance is fixed and only the mean varies.

The definition of accuracy remains effectively the same: A mechanism $\mathcal{M} : \mathbb{R}^{n\times d} \to [-1,1]^d$ is α-accurate if

$$\forall \mu \in [-1,1]^d \ \forall \sigma \in [0,\sigma_{\max}]^d \ \forall j \in [d] \quad \left| \underset{x_{1\cdots n}\sim N(\mu,\mathrm{diag}(\sigma)^2)}{\mathbb{E}}\left[\mathcal{M}(x_1,\cdots,x_n)^j\right] - \mu^j \right| \le \alpha$$

and, moreover, this statement holds when we condition on the randomness in columns other than j.

Proposition 33. *Suppose the mechanism $\mathcal{M} : \mathbb{R}^{n\times d} \to [-1,1]^d$ is α-accurate. Suppose pairs $(\mu_1,\sigma_1),\cdots,(\mu_d,\sigma_d) \in [-1,1]\times[0,\sigma_{\max}]$ are independent random variables whose distributions are all (α,γ)-strong for Gaussians. Let $x_1,\cdots x_n, z \sim N(\mu,\mathrm{diag}(\sigma)^2)$ be independent and $q = \mathcal{M}(x_1,\cdots,x_n)$. Then, for all $j \in [d]$,*

$$\mathbb{E}\left[\sum_{i\in[n]} (x_i - z)^j \, q^j\right] \ge \gamma$$

and this statement holds when we condition on the randomness in columns other than j. Moreover, if $\mathcal{M}$ is symmetric, then

$$\forall j \in [d] \ \forall i \in [n] \quad \mathbb{E}\left[(x_i - z,)^j q^j\right] \ge \frac{\gamma}{n}$$

and this statement holds when we condition on the randomness in columns other than j.

Proof. This follows from Lemma 31 and Definition 32. □

4.2.3 Completeness

Finally, having bounded the expected correlation, we can use concentration to obtain a high probability bound.

Lemma 34. *Let $\mathcal{M} : \mathbb{R}^{n\times d} \to [-1,1]^d$ be α-accurate. Assume $d > O(n^2\sigma_{\max}^2 \log(1/\delta)/\gamma^2)$. Suppose pairs $(\mu_1,\sigma_1),\cdots,(\mu_d,\sigma_d) \in [-1,1]\times[0,\sigma_{\max}]$ are independent random variables drawn from (α,γ)-strong distributions for Gaussians. Let $x_1,\cdots x_n, z \sim N(\mu,\mathrm{diag}(\sigma)^2)$ be independent and $q = \mathcal{M}(x_1,\cdots,x_n)$. Then*

$$\mathbb{P}\left[\sum_{i\in[n]} \langle x_i, q\rangle - \langle z, q\rangle < \frac{1}{2}\gamma d\right] \le \delta.$$

Moreover, if $\mathcal{M}$ is symmetric, then

$$\forall i \in [n] \quad \mathbb{P}\left[\langle x_i, q\rangle - \langle z, q\rangle < \frac{\gamma d}{2n}\right] \leq \delta.$$

We defer the proof to the appendix, as it is unenlightening and long.

Proposition 35 (Completeness)**.** *Suppose pairs $(\mu_1, \sigma_1), \cdots, (\mu_d, \sigma_d) \in [-1,1] \times [0, \sigma_{\max}]$ are independent random variables whose distributions are all (α, γ)-strong for Gaussians. Assume $d > O(n^2 \sigma_{\max}^2 \log(1/\delta)/\gamma^2)$. Suppose the mechanism $\mathcal{M} : \mathbb{R}^{n \times d} \to [-1,1]^d$ is α-accurate. Let $x_1, \cdots x_n, z \sim N(\mu, \mathrm{diag}(\sigma)^2)$ and $q = \mathcal{M}(x_1, \cdots, x_n)$. Then*

$$\mathbb{P}\left[\exists i \in [n] \quad \mathcal{A}'_{\delta,d,\sigma_{\max}}(x_i, q, z) = \mathrm{IN}\right] \geq 1 - \delta.$$

Moreover, if $\mathcal{M}$ is symmetric, then

$$\forall i \in [n] \quad \mathbb{P}\left[\mathcal{A}'_{\delta,d,\sigma_{\max}}(x_i, q, z) = \mathrm{IN}\right] \geq 1 - \delta.$$

Proof. By Lemma 34,

$$\mathbb{P}\left[\sum_{i \in [n]} \langle x_i, q\rangle - \langle z, q\rangle < \frac{1}{2}\gamma d\right] \leq \delta.$$

By assumption, $\frac{1}{2}\gamma d > n\tau' = n \cdot 2\sigma_{\max}\sqrt{d \ln(1/\delta)}$. Thus, with high probability $\langle x_i, q\rangle - \langle z, q\rangle > \tau'$ for some $i \in [n]$, as required. The second half of the proposition is similar. □

References

[BDMN05] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: The SuLQ framework. In *Symposium on Principles of Database Systems–PODS*, pages 128–138, New York, NY, USA, 2005. ACM.

[BLR08] Avrim Blum, Katrina Ligett, and Aaron Roth. A learning theory approach to non-interactive database privacy. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 609–618, New York, NY, USA, 2008. ACM.

[BRS+09] Rosemary Braun, William Rowe, Carl Schaefer, Jinghui Zhang, and Kenneth Buetow. Needles in the haystack: identifying individuals present in pooled genomic data. *PLoS genetics*, 5(10):e1000668, 2009.

[BS98] Dan Boneh and James Shaw. Collusion-secure fingerprinting for digital data. *IEEE Transactions on Information Theory*, 44(5):1897–1905, 1998.

[BUV14] Mark Bun, Jonathan Ullman, and Salil P. Vadhan. Fingerprinting codes and the price of approximate differential privacy. In *STOC*, pages 1–10. ACM, May 31 – June 3 2014.

[De12] Anindya De. Lower bounds in differential privacy. *Theory of Cryptography*, pages 321–338, 2012.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology - EUROCRYPT*, pages 486–503, St. Petersburg, Russia, 2006.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, pages 265–284. Springer, March 4-7 2006.

[DMT07] Cynthia Dwork, Frank McSherry, and Kunal Talwar. The price of privacy and the limits of LP decoding. In *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, STOC '07, pages 85–94, New York, NY, USA, 2007. ACM.

[DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *PODS*, pages 202–210. ACM, June 9-12 2003.

[DNT14] Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar. Efficient algorithms for privately releasing marginals via convex relaxations. In *Symposium on Computational Geometry–SoCG*, 2014.

[DRV10] Cynthia Dwork, Guy N. Rothblum, and Salil Vadhan. Boosting and differential privacy. In *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, FOCS '10, pages 51–60, Washington, DC, USA, 2010. IEEE Computer Society.

[DY08] Cynthia Dwork and Sergey Yekhanin. New efficient attacks on statistical disclosure control mechanisms. In *CRYPTO*, pages 469–480, 2008.

[EN14] Yaniv Erlich and Arvind Narayanan. Routes for breaching and protecting genetic privacy. *Nature Reviews Genetics*, 15(6):409–421, 2014.

[FMN13] Nadia Fawaz, S. Muthukrishnan, and Aleksandar Nikolov. Nearly optimal private convolution. In *Algorithms - ESA 2013 - 21st Annual European Symposium, Sophia Antipolis, France, September 2-4, 2013. Proceedings*, 2013.

[HR10] Moritz Hardt and Guy Rothblum. A multiplicative weights mechanism for privacy-preserving data analysis. In *Proc.* 51*st Foundations of Computer Science (FOCS)*, pages 61–70. IEEE, 2010.

[HSR+08] Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V Pearson, Dietrich A Stephan, Stanley F Nelson, and David W Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS genetics*, 4(8):e1000167, 2008.

[HT10] Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Symposium on Theory of Computing – STOC*, pages 705–714, Cambridge, MA, June 2010.

[HU14] Moritz Hardt and Jonathan Ullman. Preventing false discovery in interactive data analysis is hard. In *FOCS*. IEEE, October 19-21 2014.

[IGNC12] Hae Kyung Im, Eric R Gamazon, Dan L Nicolae, and Nancy J Cox. On sharing quantitative trait gwas results in an era of multiple-omics data and the limits of genomic privacy. *The American Journal of Human Genetics*, 90(4):591–598, 2012.

[JYW^{+}09] Kevin B Jacobs, Meredith Yeager, Sholom Wacholder, David Craig, Peter Kraft, David J Hunter, Justin Paschal, Teri A Manolio, Margaret Tucker, Robert N Hoover, , Gilles D Thomas, Stephen J Chanock, and Nilanjan Chaterjee. A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies. *Nature genetics*, 41(11):1253–1257, 2009.

[KRS13] Shiva Prasad Kasiviswanathan, Mark Rudelson, and Adam Smith. The power of linear reconstruction attacks. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013.

[KRSU10] Shiva Prasad Kasiviswanathan, Mark Rudelson, Adam Smith, and Jonathan Ullman. The price of privately releasing contingency tables and the spectra of random matrices with correlated rows. In *STOC*, pages 775–784, 2010.

[MN12] S. Muthukrishnan and Aleksandar Nikolov. Optimal private halfspace counting via discrepancy. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 1285–1292, 2012.

[NTZ13] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. *STOC*, 2013.

[RR10] Aaron Roth and Tim Roughgarden. Interactive privacy via the median mechanism. In *Proc. 42nd Symposium on Theory of Computing (STOC)*, pages 765–774. ACM, 2010.

[SOJH09] Sriram Sankararaman, Guillaume Obozinski, Michael I Jordan, and Eran Halperin. Genomic privacy and limits of individual detection in a pool. *Nature genetics*, 41(9):965–967, 2009.

[SU14] Thomas Steinke and Jonathan Ullman. Interactive fingerprinting codes and the hardness of preventing false discovery. In *COLT*, 2014.

[SU17] Thomas Steinke and Jonathan Ullman. Between pure and approximate differential privacy. *Journal of Privacy and Confidentiality*, 2017.

[Tar08] Gábor Tardos. Optimal probabilistic fingerprint codes. *J. ACM*, 55(2), 2008.

[Ull13] Jonathan Ullman. Answering $n^{2+o(1)}$ counting queries with differential privacy is hard. In *STOC*, pages 361–370. ACM, June 1-4 2013.

[VH09] Peter M Visscher and William G Hill. The limits of individual identification from sample allele frequencies: theory and statistical analysis. *PLoS genetics*, 5(10):e1000628, 2009.

[WLW^{+}09] Rui Wang, Yong Fuga Li, Xiao Feng Wang, Haixu Tang, and Xiao Yong Zhou. Learning your identity and disease from research papers: information leaks in genome wide association study. In *ACM Conference on Computer and Communications Security*, pages 534–544. ACM, 2009.

[ZPL+11] Xiaoyong Zhou, Bo Peng, Yong Fuga Li, Yangyi Chen, Haixu Tang, and XiaoFeng Wang. To release or not to release: evaluating information leaks in aggregate human-genome data. In *Computer Security–ESORICS 2011*, pages 607–627. Springer, 2011.

A Concentration Bounds

The following concentration result implies the concentration results we use in the earlier sections.

Theorem 36. *Let $X \in \mathbb{R}^{n\times d}$ be a random matrix such that the columns are independent—that is, $X^1, X^2, \cdots, X^d \in \mathbb{R}^n$ are independent random variables. Let $Y \in \mathbb{R}^d$ be a random variable that possibly depends on X. Suppose that $\underset{X}{\mathbb{E}}\left[e^{tX_{i,j}}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$, $i \in [n]$, and $j \in [d]$. Assume $\|Y\|_\infty \le \alpha$ with certainty. Let $a \in \mathbb{R}^n$. Define*

$$Z_j = (a^T X)_j Y_j \in \mathbb{R} \qquad \textit{and} \qquad Z = \sum_{j\in[d]} Z_j = a^T X Y \in \mathbb{R}.$$

Suppose $\mathbb{E}\left[Z_j \mid Z_{j+1} = z_{j+1}, \cdots, Z_d = z_d\right] \ge \gamma_j$ for all $j \in [d]$ and $z \in \mathbb{R}^d$. Let $\gamma = \sum_{j\in[d]} \gamma_j$. Then

$$\underset{Z}{\mathbb{P}}[Z < \gamma - \lambda] \le \exp\left(\frac{-\lambda^2}{16cd\alpha^2\|a\|_1^2}\right)$$

for all $\lambda > 0$.

In Lemma 11, $X_{i,j} = x_i^j - z^j$, $Y = q = M(x)$. The vector a specifies which subset of scores we are interested in (either $a = \vec{1}_n$ for the sum of all scores or $a = e_i$ for a single score).

Note that if X has bounded entries, it satisfies the condition of Theorem 36:

Lemma 37 (Hoeffding's Lemma)**.** *Let $X \in [a,b]$ be a random variable. Then*

$$\mathbb{E}\left[e^{t(X-\mathbb{E}[X])}\right] \le e^{t^2(b-a)^2/8}$$

for all $t \in \mathbb{R}$.

Likewise, if X has Gaussian entries, we can apply Theorem 36:

Lemma 38. *Let g be a standard Gaussian. Then, for all $t \in \mathbb{R}$, $\underset{g}{\mathbb{E}}\left[e^{tg}\right] = e^{t^2/2}$ and, if $0 \le t < 1/2$, $\underset{g}{\mathbb{E}}\left[e^{tg^2}\right] = 1/\sqrt{1-2t}$.*

To prove Theorem 36 we need the following lemmas.

Lemma 39 (Hölder's Inequality)**.** *Let $X_1, \cdots, X_n \in \mathbb{R}$ be (possibly dependent) random variables. Let $\alpha_*, \alpha_1, \cdots, \alpha_n \in [1,\infty]$ with $1/\alpha_* = \sum_{i\in[n]} 1/\alpha_i$. Then*

$$\underset{X}{\mathbb{E}}\left[e^{\alpha_* \sum_{i\in[n]} X_i}\right]^{1/\alpha_*} \le \prod_{i\in[n]} \underset{X_i}{\mathbb{E}}\left[e^{\alpha_i X_i}\right]^{1/\alpha_i}.$$

Lemma 40. *Let $X, Y \in \mathbb{R}$ be (possibly dependent) random variables. Suppose $|Y| \le 1$. Then*

$$\mathop{\mathbb{E}}_{X,Y}\left[e^{XY-\mathop{\mathbb{E}}_{X,Y}[XY]}\right] \le \mathop{\mathbb{E}}_{X,\xi}\left[e^{2\xi X}\right],$$

where $\xi \in \{\pm 1\}$ is uniform and independent of X and Y.

Proof. By Lemma **??**,

$$\mathop{\mathbb{E}}_{X,Y}\left[e^{XY-\mathop{\mathbb{E}}_{X,Y}[XY]}\right] \le \mathop{\mathbb{E}}_{X,Y,\xi}\left[e^{2\xi XY}\right] = \mathop{\mathbb{E}}_{X,Y,\xi}\left[e^{2\xi X|Y|}\right].$$

Define $\zeta \in \{0,1\}$ to be a "randomised rounding" of $|Y|$, namely $\mathop{\mathbb{E}}_{\zeta}[\zeta \mid Y] = |Y|$. By Jensen's inequality,

$$\mathop{\mathbb{E}}_{X,Y}\left[e^{XY-\mathop{\mathbb{E}}_{X,Y}[XY]}\right] \le \mathop{\mathbb{E}}_{X,Y,\xi}\left[e^{2\xi X|Y|}\right] = \mathop{\mathbb{E}}_{X,Y,\xi}\left[e^{\mathop{\mathbb{E}}_{\zeta}[2\xi X\zeta]}\right] \le \mathop{\mathbb{E}}_{X,Y,\xi,\zeta}\left[e^{2\xi X\zeta}\right].$$

By Jensen's inequality $e^0 = e^{\mathop{\mathbb{E}}_{\xi}[2\xi X]} \le \mathop{\mathbb{E}}_{\xi}\left[e^{2\xi X}\right]$. Thus

$$\begin{aligned}\mathop{\mathbb{E}}_{X,Y}\left[e^{XY-\mathop{\mathbb{E}}_{X,Y}[XY]}\right] &\le \mathop{\mathbb{E}}_{X,Y,\xi,\zeta}\left[e^{2\xi X\zeta}\right] = \mathop{\mathbb{E}}_{X,Y}\left[\mathop{\mathbb{P}}_{\zeta}[\zeta = 0]e^0 + \mathop{\mathbb{P}}_{\zeta}[\zeta = 1]\mathop{\mathbb{E}}_{\xi}\left[e^{2\xi X}\right]\right] \\ &\le \mathop{\mathbb{E}}_{X,Y}\left[\mathop{\mathbb{E}}_{\xi}\left[e^{2\xi X}\right]\right] = \mathop{\mathbb{E}}_{X,\xi}\left[e^{2\xi X}\right],\end{aligned}$$

as required. □

Lemma 41. *Let $X \in \mathbb{R}^n$ be a random variable. Suppose $\mathbb{E}\left[e^{tX_i}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$ and $i \in [n]$. Let $Y \in [-\alpha, \alpha]$ be a random variable that possibly depends on X . Let $a \in \mathbb{R}^n$. Define*

$$Z = a^T XY \in \mathbb{R}.$$

Then

$$\mathbb{E}\left[e^{t(Z-\mathbb{E}[Z])}\right] \le e^{4c\alpha^2t^2\|a\|_1^2}$$

for all $t \in \mathbb{R}$.

Proof. We may assume, without loss of generality, that $\|a\|_1 = 1$ and $\alpha = 1$. By assumption, $\mathbb{E}\left[e^{ta_iX_i}\right] \le e^{ct^2a_i^2}$ for all $i \in [n]$ and $t \in \mathbb{R}$. Now we apply Lemma 39 with $\alpha_i = 1/|a_i| \in [1,\infty]$ and $\alpha_* = 1$:

$$\mathbb{E}\left[e^{ta^TX}\right] = \mathbb{E}\left[e^{\alpha_* ta^TX}\right]^{1/\alpha_*} \le \prod_{i\in[n]} \mathbb{E}\left[e^{\alpha_i ta_iX_i}\right]^{1/\alpha_i} \le \prod_{i\in[n]} e^{c\alpha_i t^2 a_i^2} = e^{ct^2\|a\|_1} = e^{ct^2}$$

for all $t \in \mathbb{R}$. By Lemma 40,

$$\mathop{\mathbb{E}}_{Z}\left[e^{t(Z-\mathbb{E}[Z])}\right] = \mathop{\mathbb{E}}_{X,Y}\left[e^{ta^TXY-\mathop{\mathbb{E}}_{X,Y}[ta^TXY]}\right] \le \mathop{\mathbb{E}}_{X,\xi}\left[e^{2\xi ta^TX}\right] \le e^{4ct^2}$$

for all $t \in \mathbb{R}$, as required. □

Lemma 42. *Let $X \in \mathbb{R}^{n\times d}$ be a random variable such that the columns are independent. Suppose that $\mathbb{E}\left[e^{tX_{i,j}}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$, $i \in [n]$, and $j \in [d]$. Let $Y \in [-\alpha,\alpha]^d$ be a random variable that possibly depends on X. Let $a \in \mathbb{R}^n$. For $j \in [d]$, define*

$$Z_j = (a^T X)_j Y_j \in \mathbb{R} \qquad \text{and} \qquad \mu_j(z) = \mathbb{E}\left[Z_j \mid Z_{j+1} = z_{j+1}, \cdots, Z_d = z_d\right].$$

Let $Z = \sum_{j\in[d]} Z_j = a^T XY$ and $\mu(z) = \sum_{j\in[d]} \mu_j(z)$. Then

$$\mathbb{E}\left[e^{t(Z-\mu(Z))}\right] \le e^{4cd\alpha^2t^2\|a\|_1^2}$$

for all $t \in \mathbb{R}$.

Proof. Firstly, by Lemma 41,

$$\begin{aligned}
&\mathbb{E}\left[e^{t(Z_j-\mu_j(z))} \mid Z_{j+1} = z_{j+1}, \cdots, Z_d = z_d\right] \\
&= \mathbb{E}\left[e^{t(Z_j-\mathbb{E}[Z_j \mid Z_{j+1}=z_{j+1},\cdots,Z_d=z_d])} \mid Z_{j+1} = z_{j+1}, \cdots, Z_d = z_d\right] \\
&\le e^{4c\alpha^2t^2\|a\|_1^2}
\end{aligned}$$

for all $t \in \mathbb{R}$, $j \in [d]$, and $z \in \mathbb{R}^d$.

Now we prove by induction on $k \in [d]$ that

$$\mathbb{E}\left[e^{t\sum_{j\in[k]} Z_j-\mu_j(Z)} \mid Z_{k+1} = z_{k+1}, \cdots, Z_d = z_d\right] \le e^{4ck\alpha^2t^2\|a\|_1^2}$$

for all $t \in \mathbb{R}$ and $z \in \mathbb{R}^d$, from which the lemma follows by setting $k = d$.

The base case $k = 1$ is immediate from Lemma 41. Finally, the induction step:

$$\begin{aligned}
&\mathbb{E}\left[e^{t\sum_{j\in[k]} Z_j-\mu_j(Z)} \mid Z_{k+1} = z_{k+1}, \cdots, Z_d = z_d\right] \\
&= \sum_{z_k^*} \mathbb{P}\left[Z_k = z_k^*\right] \mathbb{E}\left[e^{t\sum_{j\in[k-1]} Z_j-\mu_j(Z)} \cdot e^{t(Z_k-\mu_k(Z))} \mid Z_k = z_k^*, Z_{k+1} = z_{k+1}, \cdots, Z_d = z_d\right] \\
&= \sum_{z_k^*} \mathbb{P}\left[Z_k = z_k^*\right] \cdot e^{t(z_k^*-\mu_k(z))} \cdot \mathbb{E}\left[e^{t\sum_{j\in[k-1]} Z_j-\mu_j(Z)} \mid Z_k = z_k^*, Z_{k+1} = z_{k+1}, \cdots, Z_d = z_d\right] \\
&\le \sum_{z_k^*} \mathbb{P}\left[Z_k = z_k^*\right] \cdot e^{t(z_k^*-\mu_k(z))} \cdot e^{4c(k-1)\alpha^2t^2\|a\|_1^2} \\
&= \mathbb{E}\left[e^{t(Z_k-\mu_k(z))} \mid Z_{k+1} = z_{k+1}, \cdots, Z_d = z_d\right] \cdot e^{4c(k-1)\alpha^2t^2\|a\|_1^2} \\
&\le e^{4ct^2\|a\|_1^2} \cdot e^{4c(k-1)t^2\|a\|_1^2},
\end{aligned}$$

as required. □

Proof of Theorem 36. The assumption that $\mathbb{E}\left[Z_j \mid Z_{j+1} = z_{j+1}, \cdots, Z_d = z_d\right] \ge \gamma_j$ for all $j \in [d]$ and $z \in \mathbb{R}^d$. Implies $\mu_j(Z) \ge \gamma_j$ with certainty. Thus it remains to show that Z is close to $\mu(Z)$.

By Markov's inequality and Lemma 42,

$$\mathbb{P}[Z-\mu(Z) > \lambda] \le \frac{\mathbb{E}\left[e^{t(Z-\mu(Z))}\right]}{e^{t\lambda}} \le \frac{e^{4cd\alpha^2t^2\|a\|_1^2}}{e^{t\lambda}}.$$

Setting $t = \frac{\lambda}{8cd\alpha^2\|a\|_1^2}$ gives

$$\mathbb{P}[Z - \mu(Z) > \lambda] \le e^{-t\lambda/2} = e^{\frac{-\lambda^2}{16cd\alpha^2\|a\|_1^2}},$$

as desired. □

A.1 Concentration of 2-Norm

Lemma 43. *Let $X \in \mathbb{R}^n$ be a product distribution. Suppose $\mathbb{E}\left[e^{tX_i}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$ and $i \in [n]$. Let $Y \in [-\alpha, \alpha]$ be a random variable that possibly depends on X. Let $a \in \mathbb{R}^n$. Define*

$$Z = a^T XY \in \mathbb{R}.$$

Then

$$\mathbb{E}\left[e^{t(Z-\mathbb{E}[Z])}\right] \le e^{4c\alpha^2 t^2\|a\|_2^2}$$

for all $t \in \mathbb{R}$.

Proof. We may assume, without loss of generality, that $\alpha = 1$. By assumption, $\mathbb{E}\left[e^{ta_iX_i}\right] \le e^{ct^2a_i^2}$ for all $i \in [n]$ and $t \in \mathbb{R}$. By independence,

$$\mathbb{E}\left[e^{ta^TX}\right] = \prod_{i\in[n]} \mathbb{E}\left[e^{ta_iX_i}\right] \le \prod_{i\in[n]} e^{ct^2a_i^2} = e^{ct^2\|a\|_2^2}$$

for all $t \in \mathbb{R}$. By Lemma 40,

$$\mathbb{E}_Z\left[e^{t(Z-\mathbb{E}[Z])}\right] = \mathbb{E}_X\left[e^{ta^TXY - \mathbb{E}_X\left[ta^TXY\right]}\right] \le \mathbb{E}_{X,\xi}\left[e^{2\xi ta^TX}\right] \le e^{4ct^2\|a\|_2^2}$$

for all $t \in \mathbb{R}$, as required. □

Lemma 44. *Let $X \in \mathbb{R}^n$ be a product distribution. Suppose $\mathbb{E}\left[e^{tX_i}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$ and $i \in [n]$. Let $Y \in [-\alpha, \alpha]$ be a random variable that possibly depends on X. Define*

$$\vec{V} = XY \in \mathbb{R}^n.$$

Then

$$\mathbb{E}\left[e^{\frac{t^2}{2}\left\|\vec{V}-\mathbb{E}\left[\vec{V}\right]\right\|_2^2}\right] \le e^{8nc\alpha^2t^2}$$

for all $t \in [-1/4\sqrt{c}\alpha, 1/4\sqrt{c}\alpha]$.

Proof. Let $g \in \mathbb{R}^n$ be a standard multivariate Gaussian and

$$Z = g^T(V - \mathbb{E}[V]) \in \mathbb{R}.$$

By Lemma 38,

$$\mathbb{E}_g\left[e^{tZ}\right] = \prod_{i\in[n]} \mathbb{E}_{g_i}\left[e^{t(V_i-\mathbb{E}[V_i])g_i}\right] = \prod_{i\in[n]} e^{t^2(\vec{V}_i-\mathbb{E}\left[\vec{V}_i\right])^2/2} = e^{t^2\left\|\vec{V}-\mathbb{E}\left[\vec{V}\right]\right\|_2^2/2}.$$

By Lemmas 43 and 38,

$$\mathop{\mathbb{E}}_{g,V}\left[e^{tZ}\right] \le \mathop{\mathbb{E}}_{g}\left[e^{4c\alpha^2t^2\|g\|_2^2}\right] = \prod_{i\in[n]}\mathop{\mathbb{E}}_{g}\left[e^{4c\alpha^2t^2g_i^2}\right] = \left(\frac{1}{\sqrt{1-2\cdot 4c\alpha^2t^2}}\right)^n,$$

assuming $0 \le 4c\alpha^2t^2 < 1/2$. Thus

$$\mathop{\mathbb{E}}_{V}\left[e^{t^2\left\|V-\mathbb{E}[V]\right\|_2^2/2}\right] \le \left(\frac{1}{\sqrt{1-8c\alpha^2t^2}}\right)^n \le e^{8nc\alpha^2t^2},$$

as $1/\sqrt{1-x} \le e^x$ for $0 \le x \le 1/2$. □

Lemma 45. *Let $X \in \mathbb{R}^{n\times d}$ be a product distribution. Suppose $\mathbb{E}\left[e^{tX_{i,j}}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$, $i \in [n]$, and $j \in [d]$. Let $Y \in [-\alpha,\alpha]^d$ be a random variable that possibly depends on X. For $j \in [d]$, define*

$$\vec{V}^j = X^jY^j \in \mathbb{R}^n \qquad \textit{and} \qquad \mu^j(\vec{v}^1,\cdots,\vec{v}^d) = \mathbb{E}\left[\vec{V}^j \mid \vec{V}^{j+1} = \vec{v}^{j+1},\cdots,\vec{V}^d = \vec{v}^d\right].$$

Let $\vec{V} = \sum_{j\in[d]}\vec{V}^j$ and $\mu(\vec{v}^1,\cdots,\vec{v}^d) = \sum_{j\in[d]}\mu^j(\vec{v}^1,\cdots,\vec{v}^d)$. Then

$$\mathbb{E}\left[e^{\frac{t^2}{2}\left\|\vec{V}-\mu(\vec{V}^1,\cdots,\vec{V}^d)\right\|_2^2}\right] \le e^{8ndc\alpha^2t^2}$$

for all $t \in [-1/4\sqrt{c}\alpha, 1/4\sqrt{c}\alpha]$.

The proof is analogous to that of Lemma 42.

Theorem 46. *Let $X \in \mathbb{R}^{n\times d}$ be a random matrix with independent entries. Suppose $\mathbb{E}\left[e^{tX_{i,j}}\right] \le e^{ct^2}$ for all $t \in \mathbb{R}$, $i \in [n]$, and $j \in [d]$. Let $Y \in [-\alpha,\alpha]^d$ be a random variable that possibly depends on X. For $j \in [d]$, define*

$$\vec{V}^j = X^jY^j \in \mathbb{R}^n.$$

Suppose that, for all $j \in [d]$ and $\vec{v}^1,\cdots,\vec{v}^d \in \mathbb{R}^n$,

$$\left\|\mathbb{E}\left[\vec{V}^j \mid \vec{V}^1 = \vec{v}^1,\cdots,\vec{V}^{j-1}\vec{v}^{j-1},\vec{V}^{j+1}\vec{v}^{j+1},\cdots,\vec{V}^d\vec{v}^d\right]\right\|_2 \le \beta_j.$$

Let $\vec{V} = \sum_{j\in[d]}\vec{V}^j$. Then

$$\mathbb{P}\left[\left\|\vec{V}\right\|_2 > \lambda + \sum_{j\in[d]}\beta_j\right] \le e^{\frac{nd}{2}-\frac{\lambda^2}{32c\alpha^2}}$$

for all $\lambda > 0$.

Proof. Let

$$\mu^j(\vec{v}^1,\cdots,\vec{v}^d) = \mathbb{E}\left[\vec{V}^j \mid \vec{V}^{j+1} = \vec{v}^{j+1},\cdots,\vec{V}^d = \vec{v}^d\right]$$

for $j \in [d]$ and

$$\mu(\vec{v}^1,\cdots,\vec{v}^d) = \sum_{j\in[d]}\mu^j(\vec{v}^1,\cdots,\vec{v}^d).$$

By assumption $\left\|\mu^j(\vec{v}^1,\cdots,\vec{v}^d)\right\|_2 \le \beta_j$ for all $j \in [d]$. Thus $\left\|\mu(\vec{v}^1,\cdots,\vec{v}^d)\right\|_2 \le \sum_{j\in[d]} \beta_j$ by the triangle inequality. By Lemma 45, $\mathbb{E}\left[e^{\frac{t^2}{2}\left\|\vec{V}-\mu(\vec{V}^1,\cdots,\vec{V}^d)\right\|_2^2}\right] \le e^{8ndc\alpha^2 t^2}$ for all $t \in [-1/4\sqrt{c}\alpha, 1/4\sqrt{c}\alpha]$. Thus, by Markov's inequality,

$$\mathbb{P}\left[\left\|\vec{V}-\mu(\vec{V}^1,\cdots,\vec{V}^d)\right\|_2 \ge \lambda\right] \le \frac{\mathbb{E}\left[e^{\frac{t^2}{2}\left\|\vec{V}-\mu(\vec{V}^1,\cdots,\vec{V}^d)\right\|_2^2}\right]}{e^{\frac{t^2}{2}\lambda^2}} \le e^{(8ndc\alpha^2-\lambda^2/2)t^2}.$$

Setting $t = 1/4\sqrt{c}\alpha$ gives the result. □

A.2 Proofs of Concentration Lemmas

Now we prove the various concentration lemmas we need.

Lemma 47 (Restating Lemma 11)**.** *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ,n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right] \ge \gamma+\alpha\cdot\xi$. Suppose the mechanism $\mathcal{M}:\{\pm1\}^{n\times d} \to [-1,1]^d$ is α-accurate. Assume $d > O(n^2\log(1/\delta)/\gamma^2)$. Let $x_1,\cdots x_n, z \sim \mathcal{P}_p$ and $q \sim \mathcal{M}(x_1,\cdots,x_n)$. Then*

$$\mathop{\mathbb{P}}_{p,x_{1\cdots n},z,\mathcal{M}}\left[\sum_{i\in[n]}(\langle x_i,q\rangle - \langle z,q\rangle) < \frac{\gamma}{2d}\right] \le \delta.$$

Moreover, if $\mathcal{M}$ is symmetric, then

$$\forall i \in [n] \quad \mathop{\mathbb{P}}_{p,x_{1\cdots n},z,\mathcal{M}}\left[\langle x_i,q\rangle - \langle z,q\rangle < \frac{\gamma d}{2n}\right] \le \delta.$$

Proof of Lemma 11. Let $a = \vec{1} \in \mathbb{R}^n$, $X_{i,j} = x_i^j - z^j$, and $Y = q = \mathcal{M}(x)$. Now we have

$$\sum_{i\in[n]}(\langle x_i,q\rangle - \langle z,q\rangle) = Z = a^T XY.$$

Lemma 37 implies $\mathbb{E}\left[e^{tX_{i,j}}\right] \le e^{t^2/2}$ for all i, j, and t. Let $Z_j = a^T X^j Y^j = \sum_{i\in[n]}(x_i^j - z^j)q^j$. Proposition 10 shows that

$$\mathbb{E}\left[Z_j \mid Z_{j+1}=z_{j+1},\cdots,Z_d=z_d\right] = \mathop{\mathbb{E}}_{p^j,x_1^j,\cdots,x_n^j,z^j}\left[\sum_{i\in[n]}\left(\langle x_i^j,q^j\rangle - \langle z^j,q^j\rangle\right)\right] \ge \gamma$$

for all $j \in [d]$ and $z \in \mathbb{R}^d$. Thus Theorem 36 shows that

$$\mathop{\mathbb{P}}_{Z}[Z < \gamma d - \lambda] \le \exp\left(\frac{-\lambda^2}{16cd\left\|a\right\|_1^2}\right)$$

for all $\lambda > 0$, where $c = 1/2$. In particular, setting $\lambda = \gamma d/2$ gives

$$\mathop{\mathbb{P}}_{Z}[Z < \gamma d/2] \le \exp\left(\frac{-(\gamma d/2)^2}{8dn^2}\right) = \exp\left(\frac{-\gamma^2 d}{32n^2}\right) \le \delta.$$

To prove the second part of the lemma, we set $a = \vec{e}_i$ instead. □

Proposition 48 (Restating Proposition 20)**.** *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal ρ is (ξ, n)-strong and satisfies $\mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2\right] \ge \gamma + \alpha\xi$. Suppose $\mathcal{M} : \{\pm1\}^{n\times d} \to [-1,1]^d$ is α-accurate. Let $d > O(\alpha^2 n^2 \log(1/\delta)/\gamma^2)$ and $m \ge 2\log(24n/\gamma)/\alpha^2$. Let $x_1, \ldots, x_n, z_0, z_1, \ldots, z_m \sim \mathcal{P}_p$. Let $q \sim \mathcal{M}(x_1, \ldots, x_n)$ Then*

$$\mathbb{P}\left[\sum_{i\in[n]}\left(\langle x_i - z, \lfloor q - w\rceil_\eta\rangle\right) < \frac{\gamma d}{2}\right] \le \delta$$

(recall $z = z_0$, $w = (1/m)\sum_{i=1}^m z_i$, and $\eta = 2\alpha$).

Proof of Proposition 20. Let $a = \vec{1}$, $X_{i,j} = (x_i^j - z^j)$, $Y_j = \lfloor q - w\rceil_\eta^j$, and $Z_j = (a^T X)_j Y_j = \sum_{i\in[n]}(x_i^j - z^j)\lfloor q - w\rceil_\eta^j$. By Lemma 19, the hypotheses of Theorem 36 are satisfied. Thus we have

$$\mathbb{P}\left[a^T X Y < d\left(\gamma - 4ne^{-\alpha^2 m/2}\right) - \lambda\right] \le e^{\frac{-\lambda^2}{8d\eta^2 n^2}}$$

for all $\lambda > 0$. Set $\lambda = \sqrt{8d\eta^2 n^2 \log(1/\delta)} \le \gamma d/6$. Now $4ne^{-\alpha^2 m/2} \le \gamma/6$, so the result follows. □

Lemma 49 (Restating Lemma 29)**.** *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal is (η, γ)-robustly strong. Let $\mathcal{M} : \{\pm1\}^{n\times d} \to [-1,1]^d$. Let $p \sim \mathcal{D}$, $x_1, \cdots x_n, z \sim \mathcal{P}_p$, and $q = \mathcal{M}(x_1, \cdots, x_n)$. Assume $d > O(n^2 \log(1/\delta)/??)$. Then*

$$\mathop{\mathbb{P}}_{p, x_{1\cdots n}, z, \mathcal{M}}\left[\sum_{i\in[n]}\langle x_i, q\rangle - \langle z, q\rangle + \frac{1}{\eta}\|q - p\|_1 \ge \frac{1}{2}\gamma d\right] \ge 1 - \delta.$$

Proof. □

Lemma 50 (Restating Lemma 34)**.** *Let $\mathcal{M} : \mathbb{R}^{n\times d} \to [-1,1]^d$ be α-accurate. Assume $d > O(n^2\sigma_{\max}^2 \log(1/\delta)/\gamma^2)$. Suppose pairs $(\mu_1, \sigma_1), \cdots, (\mu_d, \sigma_d) \in [-1,1] \times [0, \sigma_{\max}]$ are independent random variables drawn from (α, γ)-strong distributions for Gaussians. Let $x_1, \cdots x_n, z \sim N(\mu, \mathrm{diag}(\sigma)^2)$ be independent and $q = \mathcal{M}(x_1, \cdots, x_n)$. Then*

$$\mathbb{P}\left[\sum_{i\in[n]}\langle x_i, q\rangle - \langle z, q\rangle < \frac{1}{2}\gamma d\right] \le \delta.$$

Moreover, if $\mathcal{M}$ is symmetric, then

$$\forall i \in [n] \quad \mathbb{P}\left[\langle x_i, q\rangle - \langle z, q\rangle < \frac{\gamma d}{2n}\right] \le \delta.$$

Proof. As in the proof of Lemma 11, let $a = \vec{1} \in \mathbb{R}^n$, $X_{i,j} = x_i^j - \mu^j$, and $Y = q = \mathcal{M}(x)$. Now we have

$$\sum_{i\in[n]}(\langle x_i, q\rangle - \langle z, q\rangle) = Z = a^T X Y.$$

Lemma 38 implies $\mathbb{E}\left[e^{tX_{i,j}}\right] \le e^{\sigma_{\max}^2 t^2/2}$ for all i, j, and t. Let $Z_j = a^T X^j Y^j = \sum_{i\in[n]}(x_i^j - z^j)q^j$. Proposition 33 shows that

$$\mathbb{E}\left[Z_j \mid Z_{j+1} = z_{j+1}, \cdots, Z_d = z_d\right] = \mathop{\mathbb{E}}_{p^j, x_1^j, \cdots, x_n^j, z^j}\left[\sum_{i\in[n]}\left(\langle x_i^j, q^j\rangle - \langle z^j, q^j\rangle\right)\right] \ge \gamma$$

for all $j \in [d]$ and $z \in \mathbb{R}^d$. Thus Theorem 36 shows that

$$\mathop{\mathbb{P}}_{Z}[Z < \gamma d - \lambda] \leq \exp\left(\frac{-\lambda^2}{16cd\,\|a\|_1^2}\right)$$

for all $\lambda > 0$, where $c = \sigma_{\max}^2/2$. In particular, setting $\lambda = \gamma d/2$ gives

$$\mathop{\mathbb{P}}_{Z}[Z < \gamma d/2] \leq \exp\left(\frac{-(\gamma d/2)^2}{8\sigma_{\max}^2 d n^2}\right) = \exp\left(\frac{-\gamma^2 d}{32\sigma_{\max}^2 n^2}\right) \leq \delta.$$

To prove the second part of the lemma, we set $a = \vec{e}_i$ instead. □

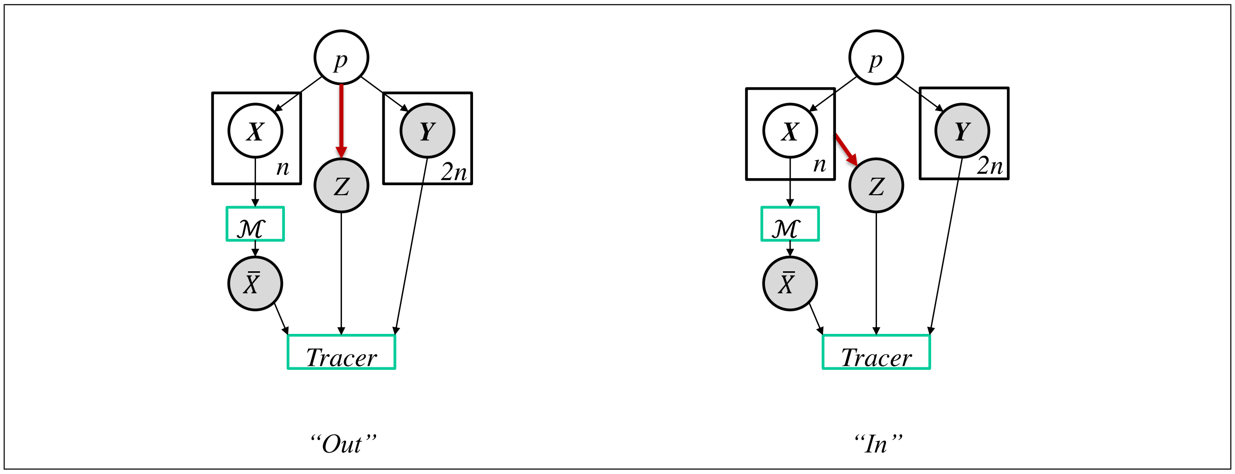

Figure 4: Graphical model depiction of the two scenarios distinguished by previous work [HSR+08, JYW+09, SOJH09]. A box with an integer k in the corner indicates k i.i.d. copies of the contents. Size $2n$ is representative of the typical size for the reference sample; [SOJH09] consider the effect of different reference sample sizes. Note that the mechanism is restricted to outputting the true sample mean $\bar{x}$.

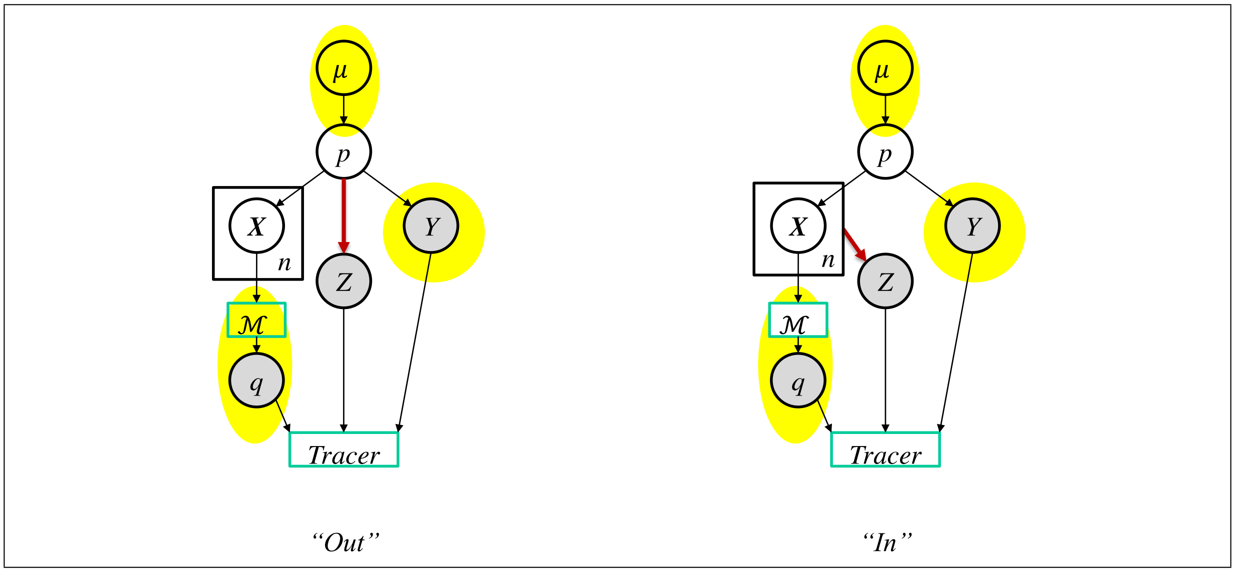

Figure 5: Graphical model depiction of the two scenarios distinguished by our algorithm. Yellow regions indicate changes relative to previous work. μ denotes the density of the distribution on the parameters p. Y denotes a single reference sample from the underlying population. q denotes the output of $\mathcal{M}(x_1,\ldots,x_n)$, which we assume is within ℓ_1 distance αd (or in somes cases ℓ_∞ distance α of the true mean $\bar{x} \in [-1,1]^d$.