Robust Traceability from Trace Amounts

Robust Traceability from Trace Amounts

Cynthia Dwork*, Adam Smith†, Thomas Steinke‡, Jonathan Ullman§ and Salil Vadhan¶
*Microsoft. Mountain View, CA, USA. dwork@microsoft.com
†Computer Science and Engineering Department.
Pennsylvania State University. State College, PA, USA. asmith@cse.psu.edu
‡John A. Paulson School of Engineering and Applied Sciences.
Harvard University. Cambridge, MA, USA. tsteinke@seas.harvard.edu
§College of Computer and Information Science.
Northeastern University. Boston, MA, USA. j.ullman@neu.edu
¶Center for Research on Computation & Society and
John A. Paulson School of Engineering and Applied Sciences.
Harvard University. Cambridge, MA, USA. salil@seas.harvard.edu

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 1-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, 2015), obviating the need for the attacker to control the exact distribution of the data.

Keywords

privacy; genomic data; fingerprinting;

I. 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

Cynthia Dwork was supported by the Simons Foundation and by the DIMACS/Simons Collaboration in Cryptography through NSF grant CNS-1523467. Part of this work was done while visiting the Simons Institute for the Theory of Computing.

Adam Smith was supported by NSF award IIS-1447700 and a Google Faculty Award. Part of this work was done while visiting Harvard University's Center for Research on Computation & Society, where he was supported by a Simons Investigator award to Salil Vadhan.

Thomas Steinke was supported by NSF grants CNS-1237235, CCF-1116616, and CCF-1420938.

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.

Salil Vadhan was supported by NSF grant CNS-1237235, a gift from Google, Inc., and a Simons Investigator grant.

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. 1-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.* [1] 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 [2].

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. Thus we ask if there is a single 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. We present and analyze such an attack.

A line of work initiated by Dinur and Nissim [3] 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 [4]–[11].

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 [12], has seen many recent developments [13]–[18]. 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 [14], [18] provide tracing attacks, based on the use of *fingerprinting codes* [19], [20], 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 [14], [18], 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 [3], [21]–[24]). 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.

A. 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 $[0,1]$).

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 1-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 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

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

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.

B. 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 $\mathcal{A} : \{\pm 1\}^d \times [-1,1]^d \times \{\pm 1\}^d \to \{\text{IN}, \text{OUT}\}$ the following holds.*

Let $\mathcal{D}$ be a product distribution on $[-1,1]^d$ such that each marginal satisfies a technical smoothness condition (Definitions 6 and 27). 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

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

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 the first input is a random element of the dataset (x_i for a random i), $\mathcal{A}$ 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 $d = \tilde{\Omega}(n^2)$ 1-way marginals with accuracy $\alpha = o(1)$, while satisfying the strong guarantee of $(o(1), 1/n^{\omega(1)})$-differential privacy [3], [21]–[24].[2] (Our attack implies that no mechanism satisfying the above conditions can be $(0.1, 1/4n)$ differentially private.) For the 1-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 [24]–[27] 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 III).

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 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 $y \in \{x_1, \ldots, x_n\}$.

[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'.

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

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

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).

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 $d = O(n^2\sigma^2_{\max}\log(1/\delta))$.

C. Description of The Attack

Like the attacks in previous tracing work for the genomic setting [1], [2], [28]–[30] and in the fingerprinting setting [17], [20], 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 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 feature vector p, or the distribution $\mathcal{D}$ on p.

D. Comparison with Previous Work

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

- Work in the genomics community [1], [2], [28], [29], [31] 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 [2]). 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.* [32] 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 [14], [18] 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 [1], [29], [32], [33]; see [34] for a survey. Particularly exciting (or troubling) is the work of Wang *et al.* [33] 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*.

II. Tracing with a Single Reference Sample

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

A. 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*

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

Proof: We can view p and q as fixed. Since y and z are identically distributed, $\mathbb{E}\left[\langle y, q\rangle - \langle z, q\rangle\right] = 0$. 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 $\max\{||y||_\infty, ||z||_\infty\} \cdot ||q||_\infty \leq 1$. Thus, by a Chernoff bound,

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

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.*

B. 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)**.** *Let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ be a (possibly randomized) mechanism. Fix any $p \in [-1,1]^d$. Consider the following experiment. Let $x_1, \ldots, x_n \sim \mathcal{P}_p$ and then let $q \sim \mathcal{M}(x_1, \ldots, x_n)$. We say that $\mathcal{M}$ is* α-strongly accurate *if for every $j \in [d]$,*

$$\left|\mathbb{E}\left[q^j\right] - p^j\right| \leq \alpha$$

and, moreover, this statement holds even when we condition on all the randomness in columns other than j. That is, accuracy must hold when we condition on any values of $\{x_i^{-j}\}_{i=1,\ldots,n}$ and q^{-j} and the randomness is taken only over the remaining variables.

Note that if $\mathcal{M}$ satisfies $\left|\mathcal{M}(x)^j - \frac{1}{n}\sum_{i\in[n]} x_i^j\right| \leq \alpha$ for all $x \in \{\pm 1\}^{n\times d}$ and $j \in [d]$, then $\mathcal{M}$ satisfies α-accuracy. In Section IV-A, we discuss mechanisms that satisfy a weaker "ℓ_1" accuracy condition.

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

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

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 $\mathbb{E}\left[\sum_{i\in[n]} x_i^j q^j - z^j q^j\right] \geq Cn\tau/d$ for each $j \in [d]$. We now focus on a fixed $j \in [d]$ and, for clarity, omit the superscript. The following lemmas yield a proof of this statement.

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 $x_1, \cdots x_n \in \{\pm 1\}^n$ are drawn independently with $x_i \sim \rho$ for each $i \in [n]$.

Lemma 5. *Let $f : \{\pm 1\}^n \to \mathbb{R}$. Define $g : [-1,1] \to \mathbb{R}$ by*

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

Then

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

This result is similar to [17, Lemma 2.11]. (It can be viewed as a rescaling of said lemma.)

Proof: Since $x^2 = 1$ for $x \in \{\pm 1\}$, 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 $x \in \{\pm 1\}$ 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 $x \in \{\pm 1\}^n$ and $p \in (-1, 1)$. Sampling $x \sim p$ samples each $x \in \{\pm 1\}$ with probability $\frac{1+xp}{2}$. Thus sampling $x_{1\cdots n} \sim p$, samples each $x \in \{\pm 1\}^n$ with probability $\prod_{i\in[n]} \frac{1+x_i p}{2}$.

Now we can write

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

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. ∎

The following definition is the technical smoothness condition we need the marginals of the distribution to satisfy.

Definition 6 (Strong Distribution)**.** *A probability distribution ρ on $[-1, 1]$ is (α, γ)-*strong *if*

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

for all polynomials $g : [-1, 1] \to [-1, 1]$ satisfying $|g(p) - p| \leq \alpha$ for all $p \in [-1, 1]$.

We give some meaning to this definition in Section II-D. 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 $(1/3, 1/3)$-strong.

Lemma 7. *Let $f : \{\pm 1\}^n \to [-1, 1]$. Define $g : [-1, 1] \to [-1, 1]$ by*

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

Assume that $|g(p) - p| \leq \alpha$ for all $p \in [-1, 1]$. Let ρ be a (α, γ)-strong probability distribution. Then

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

Proof: By Lemma 5 and Definition 6,

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

∎

We now make an observation that will allow the construction of a high-power attack. 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 Lemma 7 can be altered to

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

Formally, we have the following definition and Lemma.

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

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

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

for all $k \in [n]$.

Proof: By Definition 8,

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

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. ∎

Combining the above Lemmas shows that, if p is drawn from a strong distribution and $\mathcal{M}$ is α-accurate, then we have large expected score.

Proposition 10. *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal is (α, γ)-strong. 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

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

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 $p^{-j}, \{x_i^{-j}\}_{i=1,\ldots,n}, z^{-j}, q^{-j}$ and the randomness is only over the remaining variables.

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

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

and hence

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

Proof: We view z^{-j}, q^{-j}, x_i^{-j} as fixed and we average over the coins 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, $|\mathop{\mathbb{E}}_{x_{1\cdots,n}\sim p}[f(x)] - p| \leq \alpha$ for all $p \in [-1, 1]$. The result now follows from Lemmata 7 and 9. ∎

C. Completeness Analysis

Now that we have shown that $\mathbb{E}\left[\sum_{i\in[n]}\left(\langle x_i, q\rangle - \langle z, q\rangle\right)\right]$ 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 is (α,γ)-strong. 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*

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

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

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

The formal proof of this Lemma is quite involved, but unenlightening. Thus we defer it to the full version of this work and give a proof sketch here instead.

Proof: [Proof Sketch] Write

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

We have $\mathbb{E}[A_j] \ge \gamma$ for all $j \in [d]$. Suppose the A_j random variables were independent. Then we could apply a Chernoff bound. Using $|A_j| \le 2n$, gives

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

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 $\mathbb{E}[A_j] \ge \gamma$ even if we condition on $A_1, \cdots, A_{j-1}, A_{j+1}, \cdots, A_d$. Namely

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

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 is (α,γ)-strong. Assume $d > O(n^2\log(1/\delta)/\gamma^2)$. 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 $q = \mathcal{M}(x_1, \cdots, x_n)$. Then*

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

Proof: By Lemma 11, $\sum_{i\in[n]}\left(\langle x_i, q\rangle - \langle z, q\rangle\right) \ge \frac{\gamma}{2d} > n\cdot\tau = n\cdot 2\sqrt{d\log(1/\delta)}$ 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 is (α,γ)-strong. Assume $d > O(n^2\log(1/\delta)/\gamma^2)$. Suppose the mechanism $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ 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*

$$\forall i \in [n] \quad \mathop{\mathbb{P}}_{p,x_1,\cdots,x_n,z,\mathcal{M}}[\mathcal{A}_{\delta,d}(x_i, q, z) = \text{IN}] \ge 1 - \delta.$$

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

D. Interpreting Strong Distributions

The notion of strong distributions (Definition 6) 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.

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^*. Let[3]

$$g(p) = \begin{cases} p - \alpha & p > p^* + \alpha \\ p^* & |p - p^*| \leq \alpha \\ p + \alpha & p < p^* - \alpha \end{cases}.$$

Then $\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] = 0$.[4] Thus a point mass is not (α, γ)-strong for any $\alpha, \gamma > 0$.

This g corresponds to a mechanism $\mathcal{M}$ that knows p^* and outputs p^* instead of $\frac{1}{n}\sum_{i\in[n]} x_i$ (unless $\left|\frac{1}{n}\sum_{i\in[n]} x_i - p^*\right| > \alpha$, which is unlikely). Since this mechanism's answers don't depend on its input, we cannot hope to trace the members of the dataset.

(ii) Example (i) can be generalized: Any distribution supported on an interval of length α is not (α, γ)-strong for any $\alpha, \gamma > 0$.

(iii) Suppose ρ is supported on p^* and p^{**} with $p^* < p^{**} - \alpha$. We can construct a piecewise linear $g : \mathbb{R} \to \mathbb{R}$ with $g'(p) = 0$ if $|p - p^*| \leq \alpha/3$ or $|p - p^{**}| \leq \alpha/3$ and $g(p^*) = p^*$ and $g(p^{**}) = p^{**}$. This gives $|g(p) - p| \leq \alpha$ for all p and $\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] = 0$. Thus this distribution is not (α, γ)-strong for any $\alpha, \gamma > 0$.

This g 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 generalized to any distribution supported on points that are separated by distance α. This can be generalized further to distributions supporeted on many intervals of size α that are also separated by distance α. These distributions amount to separated narrow clumps of probability mass.

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

In the special case of $\alpha = 0$, which corresponds to $\mathcal{M}$ giving unbiased answers, any distribution is $(0, 1 - \mathop{\mathbb{E}}_{p\sim\rho}\left[p^2\right])$-strong.

In general, the "ideal" strong distribution (*cf.* [20]) is as follows. Let ρ have support $[-a, a]$ and probability density function $\rho(p) \propto 1/(1-p^2)$. Then, for $g : [-1, 1] \to [-1, 1]$ satisfying $|g(p) - p| \leq \alpha$ for $p \in \{\pm a\}$,

$$\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] = \int_{-a}^{a} g'(p)(1-p^2)\rho(p)\mathrm{d}p = \frac{g(a) - g(-a)}{\int_{-a}^{a}(1-p^2)^{-1}\mathrm{d}p} \geq \frac{2a - 2\alpha}{\log(1+a) - \log(1-a)}.$$

For example, setting $a = 0.85$ makes ρ $(0.5, 0.27)$-strong.

It is unreasonable to expect that this exact distribution will arise in nature. However, we can show that a reasonably large class of distributions (including the uniform distribution) are all strong.

First we prove a technical lemma that reinterprets the strong condition.

[3]This g is not continuously differentiable, but may be approximated arbitrarily well by a continuously differentiable $\tilde{g}$.

[4]In fact, we can construct g with $\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] \ll 0$.

Lemma 14. *Let $\rho : [a,b] \to \mathbb{R}$ be a continuously differentiable probability density function. Let $g : [-1,1] \to [-1,1]$ be continuously differentiable. Then*

$$\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] = 1 - \mathop{\mathbb{E}}_{p\sim\rho}\left[p^2\right] + (g(b)-b)(1-b^2)\rho(b) - (g(a)-a)(1-a^2)\rho(a) + 2\mathop{\mathbb{E}}_{p\sim\rho}\left[(g(p)-p)p\right] - \int_a^b (g(p)-p)(1-p^2)\rho'(p)\mathrm{d}p. \tag{2}$$

In particular, if $[a,b] = [-1,1]$ or $\rho(b) = 0 = \rho(a)$, then

$$\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] = 1 - \mathop{\mathbb{E}}_{p\sim\rho}\left[p^2\right] + 2\mathop{\mathbb{E}}_{p\sim\rho}\left[(g(p)-p)p\right] - \int_a^b (g(p)-p)(1-p^2)\rho'(p)\mathrm{d}p. \tag{3}$$

Proof: We have

$$\mathop{\mathbb{E}}_{p\sim\rho}\left[g'(p)(1-p^2)\right] = \mathop{\mathbb{E}}_{p\sim\rho}\left[1-p^2+(g'(p)-1)(1-p^2)\right] = 1 - \mathop{\mathbb{E}}_{p\sim\rho}\left[p^2\right] + \int_a^b (g'(p)-1)(1-p^2)\rho(p)\mathrm{d}p. \tag{4}$$

Integration by parts gives

$$\int (g'(p)-1)(1-p^2)\rho(p)\mathrm{d}p = (g(p)-p)(1-p^2)\rho(p) - \int (g(p)-p)(-2p\rho(p) + (1-p^2)\rho'(p))\mathrm{d}p.$$

The fundamental theorem of calculus gives

$$\int_a^b (g'(p)-1)(1-p^2)\rho(p)\mathrm{d}p = (g(b)-b)(1-b^2)\rho(b) - (g(a)-a)(1-a^2)\rho(a) + 2\int_a^b (g(p)-p)p\rho(p)\mathrm{d}p - \int_a^b (g(p)-p)(1-p^2)\rho'(p)\mathrm{d}p. \tag{5}$$

Combining (4) with (5) gives (2). If $[a,b] = [-1,1]$ or $\rho(b) = 0 = \rho(a)$, then $(1-b^2)\rho(b) = 0 = (1-a^2)\rho(a)$, which implies (3). ∎

Now we can use Lemma 14 to show that various distributions are strong:

Corollary 15. *Let $\rho : [-1,1] \to \mathbb{R}$ be a continuously differentiable probability density function. Then ρ is (α,γ)-strong for all α and*

$$\gamma = 1 - \mathop{\mathbb{E}}_{p\sim\rho}\left[p^2\right] - 2\alpha\mathop{\mathbb{E}}_{p\sim\rho}\left[|p|\right] - \alpha\int_{-1}^{1}\left|\rho'(p)\right|(1-p^2)\mathrm{d}p.$$

Corollary 16. *Let $\rho : [a,b] \to \mathbb{R}$ be a continuously differentiable probability density function. Then ρ is (α,γ)-strong for all α and*

$$\gamma = 1 - \mathop{\mathbb{E}}_{p\sim\rho}\left[p^2\right] - \alpha(\rho(b)+\rho(a)) - 2\alpha\mathop{\mathbb{E}}_{p\sim\rho}\left[|p|\right] - \alpha\int_{a}^{b}\left|\rho'(p)\right|(1-p^2)\mathrm{d}p.$$

Corollaries 15 and 16 give sufficient conditions for a distribution to be strong. Intuitively, Corollary 15 says that a smooth distribution (meaning $\int_{-1}^{1}|\rho'(p)|(1-p^2)\mathrm{d}p = O(1)$) is (α,γ)-strong for $\gamma = 1 - \mathbb{E}\left[p^2\right] - O(\alpha)$.

Now we can give examples of strong distributions:

- The uniform distribution on $[-1,1]$ is $(\alpha, 2/3-\alpha)$-strong for all $\alpha \le 2/3$.
- The uniform distribution on $[a,b]$ is (α,γ)-strong for
$$\gamma = 1 - \frac{b^2+ab+a^2}{3} - \frac{2\alpha}{b-a} - 2\alpha \ge \frac{2}{3} - \frac{2}{3}(b-a) - \frac{2\alpha}{b-a} - 2\alpha.$$
- 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 (α,γ)-strong for
$$\gamma = \frac{4uv}{(u+v+1)(u+v)} - 2\alpha\sqrt{1 - \frac{4uv}{(u+v+1)(u+v)}} - 2\alpha\frac{v|u-1| + u|v-1|}{u+v}.$$

III. 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 summarize our results in the following informal theorem, which effectively generalizes 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 (Definitions 6 and 27). 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

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

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

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.

$$\mathcal{A}^*_{\delta,\alpha,d,m}(y, q, \vec{z})$$

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 $[-\eta, \eta]$.
4. Compute
$$\langle y - z, \lfloor q - w \rceil_\eta \rangle = \sum_{j \in [d]} (y^j - z^j) \cdot \lfloor q^j - w^j \rceil_\eta .$$
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**

A. 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,

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

This completes the proof. ■

B. Correlation Analysis

We have the following proposition, analogous to Proposition 10 in Section II-B.

Before diving into the analysis, we need to slightly strengthen our definition of accuracy. Instead of assuming merely that the expected error of the mechanism is α, we want to assume that the mechanism's error is bounded by α with high probability.

Definition 19 (Strong Accuracy)**.** *Let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ be a (possibly randomized) mechanism. Fix any $p \in [-1,1]^d$. Consider the following experiment. Let $x_1, \dots, x_n \sim \mathcal{P}_p$ and then let $q \sim \mathcal{M}(x_1, \dots, x_n)$. We say that $\mathcal{M}$ is* (α, β)-strongly accurate *if for every $j \in [d]$,*

$$\mathbb{P}\left[|q^j - p^j| > \alpha\right] \le \beta,$$

and, moreover, this statement holds even when we condition on all the randomness in columns other than j. That is, accuracy must hold when we condition on any values of $\{x_i^{-j}\}_{i=1,\dots,n}$ and q^{-j} and the randomness is taken only over the remaining variables.

Note that if $\mathcal{M}$ satisfies $\left|\mathcal{M}(x)^j - (1/n)\sum_{i\in[n]} x_i^j\right| \le \alpha$ for all $x \in \{\pm 1\}^{n\times d}$ and $j \in [d]$, then $\mathcal{M}$ satisfies $\left(\alpha + \lambda, e^{-\lambda^2 n/2}\right)$-strong accuracy for all $\lambda \ge 0$. That is, if $\mathcal{M}$ is close to the *empirical mean* of its input samples and n is large, then $\mathcal{M}$ is also close to the population mean p.

Lemma 20. *Let $\mathcal{M} : \{\pm 1\}^{n\times d} \to [-1,1]^d$ be (α, β)-strongly accurate, let $\eta = 2\alpha$, and let the distribution $\mathcal{D}$ be a product distribution where every marginal is (α, γ)-strong. Consider the following experiment. Let $p \sim \mathcal{D}$, let $x_1, \dots, x_n, z_0, z_1, \dots, z_m \sim \mathcal{P}_p$, and $q \sim \mathcal{M}(x_1, \dots, x_n)$. Then for every $j \in [d]$,*

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

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}, $\{x_i^{-j}\}_{i=1,\dots,n}$, $\{z_i^{-j}\}_{j=0,1,\dots,m}$, and q^{-j} and the randomness is taken only over the remaining variables.

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

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

So it remains to show that

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

Since $\left|\sum_{i\in[n]} (x_i^j - z^j) \cdot (q^j - w^j - \left\lfloor q^j - w^j \right\rceil_\eta)\right| \le 4n$ 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 $|q^j - w^j| \le \eta$, it suffices to show that $\mathbb{P}\left[|q^j - w^j| > \eta\right] \le \beta + e^{-\alpha^2 m/2}$. By strong accuracy, we have

$\mathbb{P}\left[|q^j - p^j| > \alpha\right] \leq \beta$, 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 21. *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal is (α, γ)-strong. Suppose $\mathcal{M} : \{\pm 1\}^{n \times d} \to [-1, 1]^d$ is (α, β)-strongly accurate for $\beta \leq \gamma/24n$. 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 21 is analogous to that of Lemma 11 and is deferred to the appendix.

Proposition 21 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 22. *Let $X_1, \cdots, X_n \in \mathbb{R}$ be independent random variables such that $\mathbb{E}[X_i] = 0$ and $\mathbb{E}\left[X_i^2\right] \leq 1$ for every $i \in [n]$. Let $Y \in \mathbb{R}$ be another (not necessarily independent) random variable. Then*

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

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

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

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

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

Rearranging gives

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

as required. ∎

Lemma 23. *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 $j \in [d]$,*

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

(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 apple Lemma 22 with $X_i = x_i^j - z^j$ and $Y = \lfloor q^j - w^j \rceil_\eta$. ∎

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 the full version of this work. Using this concentration bound we can prove the following.

Proposition 24. *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 23, 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 our concentration result from the full version of this work, 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 21 with Proposition 24, 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 25. *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 26 (Completeness with a Large Reference Sample)**.** *Suppose the distribution $\mathcal{D}$ is a product distribution in which each marginal is (α,γ)-strong. Suppose$\mathcal{M}:\{\pm1\}^{n\times d}\to[-1,1]^d$ is (α,β)-strongly accurate for $\beta\le\gamma/24n$. Let $d>O(\alpha^2n^2\log(1/\delta)/\gamma^2)$ and $m\ge 2\log(24n/\gamma)/\alpha^2$. 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*

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

Proof: By Proposition 21, with probability at least $1-\delta$,

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

By Proposition 24, with probability at least $1-\delta$,

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

By a union bound, both of these events occur with probability at least $1-2\delta$. Assuming they both occur, Lemma 25 implies

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

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

IV. Extensions

A. Robustness: Mechanisms with ℓ_1-Bounded Error

We have taken $\mathcal{M}$ being accurate to mean $\left|\left|\mathbb{E}[q]-p\right|\right|_\infty\le\alpha$ 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, $\left|\left|\mathbb{E}[q]-p\right|\right|_1\le\alpha d$ – a very weak average-case error guarantee.

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

Definition 27 (Robustly Strong Distribution)**.** *A probability distribution ρ on $[-1,1]$ is* (η,γ)-robustly strong *if*

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

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 28 (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 $\mathcal{M}:\{\pm1\}^{n\times d}\to[-1,1]^d$. Let $p\sim\mathcal{D}$, $x_1,\cdots x_n,z\sim\mathcal{P}_p$, and $q=\mathcal{M}(x_1,\cdots,x_n)$. Then*

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

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

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**

B. 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.

Verifying soundness of our attack is again straightforward.

Proposition 29 (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.*[5] *Then*

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

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 $2\sum_{j\in[d]} \sigma_j^2 q_j^2 \le 2d\sigma_{\max}^2$, we have

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

as required. ∎

The relevant notion of smoothness is now the following.

Definition 30. *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*

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

Proposition 31 (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*

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

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

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

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

REFERENCES

[1] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. V. Pearson, D. A. Stephan, S. F. Nelson, and D. W. Craig, "Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays," *PLoS genetics*, vol. 4, no. 8, p. e1000167, 2008.

[2] S. Sankararaman, G. Obozinski, M. I. Jordan, and E. Halperin, "Genomic privacy and limits of individual detection in a pool," *Nature genetics*, vol. 41, no. 9, pp. 965–967, 2009.

[3] I. Dinur and K. Nissim, "Revealing information while preserving privacy," in *PODS*. ACM, June 9-12 2003, pp. 202–210.

[4] C. Dwork, F. McSherry, and K. Talwar, "The price of privacy and the limits of LP decoding," in *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing*, ser. STOC '07. New York, NY, USA: ACM, 2007, pp. 85–94.

[5] C. Dwork and S. Yekhanin, "New efficient attacks on statistical disclosure control mechanisms," in *CRYPTO*, 2008, pp. 469–480.

[6] S. P. Kasiviswanathan, M. Rudelson, A. Smith, and J. Ullman, "The price of privately releasing contingency tables and the spectra of random matrices with correlated rows," in *STOC*, 2010, pp. 775–784.

[7] A. De, "Lower bounds in differential privacy," *Theory of Cryptography*, pp. 321–338, 2012.

[8] S. P. Kasiviswanathan, M. Rudelson, and A. Smith, "The power of linear reconstruction attacks," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013.

[9] S. Muthukrishnan and A. 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*, 2012, pp. 1285–1292.

[10] N. Fawaz, S. Muthukrishnan, and A. Nikolov, "Nearly optimal private convolution," in *Algorithms - ESA 2013 - 21st Annual European Symposium, Sophia Antipolis, France, September 2-4, 2013. Proceedings*, 2013.

[11] A. Nikolov, K. Talwar, and L. Zhang, "The geometry of differential privacy: the sparse and approximate cases," *STOC*, 2013.

[12] M. Hardt and K. Talwar, "On the geometry of differential privacy," in *Symposium on Theory of Computing – STOC*, Cambridge, MA, June 2010, pp. 705–714.

[13] J. Ullman, "Answering $n^{2+o(1)}$ counting queries with differential privacy is hard," in *STOC*. ACM, June 1-4 2013, pp. 361–370.

[14] M. Bun, J. Ullman, and S. P. Vadhan, "Fingerprinting codes and the price of approximate differential privacy," in *STOC*. ACM, May 31 – June 3 2014, pp. 1–10.

[15] C. Dwork, A. Nikolov, and K. Talwar, "Efficient algorithms for privately releasing marginals via convex relaxations," in *Symposium on Computational Geometry–SoCG*, 2014.

[16] M. Hardt and J. Ullman, "Preventing false discovery in interactive data analysis is hard," in *FOCS*. IEEE, October 19-21 2014.

[17] T. Steinke and J. Ullman, "Interactive fingerprinting codes and the hardness of preventing false discovery," in *COLT*, 2014.

[18] ——, "Between pure and approximate differential privacy," *arXiv preprint arXiv:1501.06095*, 2015.

[19] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Transactions on Information Theory*, vol. 44, no. 5, pp. 1897–1905, 1998.

[20] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM*, vol. 55, no. 2, 2008.

[21] A. Blum, C. Dwork, F. McSherry, and K. Nissim, "Practical privacy: The SuLQ framework," in *Symposium on Principles of Database Systems–PODS*. New York, NY, USA: ACM, 2005, pp. 128–138.

[22] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *TCC*. Springer, March 4-7 2006, pp. 265–284.

[23] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Advances in Cryptology - EUROCRYPT*, St. Petersburg, Russia, 2006, pp. 486–503.

[24] C. Dwork, G. N. Rothblum, and S. Vadhan, "Boosting and differential privacy," in *Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, ser. FOCS '10. Washington, DC, USA: IEEE Computer Society, 2010, pp. 51–60.

[25] A. Blum, K. Ligett, and A. Roth, "A learning theory approach to non-interactive database privacy," in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, ser. STOC '08. New York, NY, USA: ACM, 2008, pp. 609–618.

[26] M. Hardt and G. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis," in *Proc. 51st Foundations of Computer Science (FOCS)*. IEEE, 2010, pp. 61–70.

[27] A. Roth and T. Roughgarden, "Interactive privacy via the median mechanism," in *Proc. 42nd Symposium on Theory of Computing (STOC)*. ACM, 2010, pp. 765–774.

[28] R. Braun, W. Rowe, C. Schaefer, J. Zhang, and K. Buetow, "Needles in the haystack: identifying individuals present in pooled genomic data," *PLoS genetics*, vol. 5, no. 10, p. e1000668, 2009.

[29] K. B. Jacobs, M. Yeager, S. Wacholder, D. Craig, P. Kraft, D. J. Hunter, J. Paschal, T. A. Manolio, M. Tucker, R. N. Hoover, , G. D. Thomas, S. J. Chanock, and N. Chaterjee, "A new statistic and its power to infer membership in a genome-wide association study using genotype frequencies," *Nature genetics*, vol. 41, no. 11, pp. 1253–1257, 2009.

[30] X. Zhou, B. Peng, Y. F. Li, Y. Chen, H. Tang, and X. Wang, "To release or not to release: evaluating information leaks in aggregate human-genome data," in *Computer Security–ESORICS 2011*. Springer, 2011, pp. 607–627.

[31] P. M. Visscher and W. G. Hill, "The limits of individual identification from sample allele frequencies: theory and statistical analysis," *PLoS genetics*, vol. 5, no. 10, p. e1000628, 2009.

[32] H. K. Im, E. R. Gamazon, D. L. Nicolae, and N. 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*, vol. 90, no. 4, pp. 591–598, 2012.

[33] R. Wang, Y. F. Li, X. F. Wang, H. Tang, and X. 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*. ACM, 2009, pp. 534–544.

[34] Y. Erlich and A. Narayanan, "Routes for breaching and protecting genetic privacy," *Nature Reviews Genetics*, vol. 15, no. 6, pp. 409–421, 2014.