Randomized Response Schemes, Privacy and Usefulness[*]

Francesco Aldà
Horst Görtz Institute for IT Security
and Faculty of Mathematics
Ruhr-Universität Bochum, Germany
francesco.alda@rub.de

Hans Ulrich Simon
Horst Görtz Institute for IT Security
and Faculty of Mathematics
Ruhr-Universität Bochum, Germany
hans.simon@rub.de

ABSTRACT

We define the notions of weak and strong usefulness and investigate the question whether a randomized response scheme can combine two conflicting goals, namely being weakly (or even strongly) useful and, at the same time, providing ε-differential privacy. We prove the following results. First, if a class $\mathcal{F}$ cannot be weakly SQ-learned under the uniform distribution, then ε-differential privacy rules out weak usefulness. This extends a result from [6] that was concerned with the class of parities. Second, for a broad variety of classes $\mathcal{F}$ that actually can be weakly SQ-learned under the uniform distribution, we design a randomized response scheme that provides ε-differential privacy and strong usefulness.

Categories and Subject Descriptors

F.2.m [**Analysis of Algorithms and Problem Complexity**]: Miscellaneous; K.4.1 [**Computers and Society**]: Public Policy Issues—*Privacy*

General Terms

Theory

Keywords

SQ-learning; Margin Complexity; Privacy

1. INTRODUCTION

The perspectives of private data analysis and learning theory are, in some sense, dual to each other, and both areas are closely related. In learning theory, one has a sample (a kind of database) and would like to infer information about an unknown target concept. In private data analysis, one has only indirect access to a database via queries. Counting queries, in particular, are represented by predicates (like binary concepts in learning theory). An analyst of a database D chooses a query f (roughly corresponding to a concept) and wants to "learn" something about the unknown data in D. Typically, she is not interested in the precise or approximate reconstruction of D but wants to get some piece of aggregated information only (like, for instance, the fraction of instances in D satisfying the predicate f in case of a counting query). With direct access to D, this would be child's play. However, if D contains sensitive information, the following obstacles will make the task of learning something about the data more challenging:

- We would like the database access mechanism to provide "ε-differential privacy" [6] (a very well-established notion by now).
- We would like to use very special database access mechanisms named "randomized response schemes" [6]. They have the advantage of not relying on trust in the database manager.

This paper investigates to what extent randomized response schemes $\mathcal{M}$ can combine ε-differential privacy with the conflicting goal of providing useful answers to counting queries taken from a (known) class $\mathcal{F}$. In order to provide ε-differential privacy, the database D is transformed into a "noisy database" $\hat{D} = \mathcal{M}(D)$. The crucial question is whether one can still extract useful information from $\hat{D}$. This general problem is addressed by a series of prior works (see for example [4, 10]), which focus on a *worst-case scenario*, namely asking that, for *all* possible choices of the database D, the answers provided by the scheme are close (in probability) to the "correct" ones. It turns out that the family of queries for which this goal can be achieved is quite limited (see Section 2.4 for further details). In this paper, we relax the requirement on the usefulness (thereby avoiding any trivialization of the problem itself), by analyzing a "minimum" goal which every private randomized response scheme should accomplish. Specifically, we assume that "positive" instances (satisfying the predicate underlying the counting query) entering a database D are chosen uniformly at random from the set of all positive instances in the universe. A similar remark applies to "negative" instances. An extremely weak notion of "usefulness" is obtained when we merely want to be able to distinguish a database consisting exclusively of negative instances from a database consisting exclusively of positive ones (so that the correct answer to the query would be either 0 or 1). In [6], it is shown that, for the query class of parities, even this (extremely) weak criterion cannot

[*]The research was supported by the DFG Research Training Group GRK 1817/1

AISec'14, November 7, 2014, Scottsdale, Arizona, USA.

ACM 978-1-4503-2953-1/14/11 ...$15.00.
http://dx.doi.org/10.1145/2666652.2666654.

be satisfied by randomized response schemes that provide ε-differential privacy (unless the database becomes exponentially large). In this paper, we show the following results. First, the (just mentioned) negative result from [6] can be extended to any class that is not weakly SQ-learnable under the uniform distribution (which includes the class of parities as a special case). Second, for a broad variety of classes $\mathcal{F}$ that actually are weakly SQ-learnable under the uniform distribution, we design a randomized response scheme that provides ε-differential privacy and, at the same time, meets a quite strong criterion of usefulness for every $f \in \mathcal{F}$, namely allowing to infer (in probability) from $\hat{D}$ an approximation $\hat{\omega}$ of the true fraction ω of instances in D satisfying f.

We would like to stress that SQ-learnability is a quite influential model with rich relations to other concepts in machine learning theory like, for instance, margin complexity or evolvability [20, 7]. The results in this paper (and previous ones [14, 4, 10]) show that these concepts are of high relevance in the field of Differential Privacy too, so as to establish a strong connection between these two fields.

2. DEFINITIONS, FACTS, AND MAIN RESULTS

For a set S, $\mathbf{1}_S(x) \in \{0,1\}$ denotes the characteristic function with value 1 iff $x \in S$. For a vector v with N components, $\mathrm{diag}(v)$ denotes the $(N \times N)$-matrix with the components of v on the main diagonal and zeros elsewhere. A *probability vector* p is a vector with non-negative components that sum up to 1. The *statistical difference* between two probability measures P, Q (sometimes called the *total variation distance*) is given by

$$\mathrm{SD}(P,Q) = \frac{1}{2}\|P-Q\|_1 = \frac{1}{2}\sum_x |P(x)-Q(x)| \enspace ,$$

where, in the continuous case, the sum must be replaced by an integral. The following is known:

Lemma 1 ([17]). *Given a sample point x drawn according to P or Q with probability $1/2$, respectively, the task of deciding whether x was sampled according to P or Q has Bayes error $(1-\mathrm{SD}(P,Q))/2$.*

As usual, $\|v\|_2$ denotes the Euclidean norm of the vector v. For a matrix A, its *spectral norm*, denoted $\|\!|A|\!\|_2$, and its *Frobenius norm*, denoted $\|A\|_2$, are given by

$$\|\!|A|\!\|_2 = \max_{\|v\|_2=1} \|Av\|_2 \text{ and } \|A\|_2 = \left(\sum_{i,j} A_{i,j}^2\right)^{1/2} \enspace . \tag{1}$$

It is well known that

$$\begin{aligned} &\|\!|A|\!\|_2 \le \|A\|_2 \text{ and} \\ &\|\!|A^\top|\!\|_2^2 = \|\!|A|\!\|_2^2 = \max\{v^\top A^\top A v : \|v\|_2 = 1\} \enspace . \end{aligned} \tag{2}$$

2.1 Private Data Analysis

Let $\mathcal{X}$ be a finite universe which can be thought of as a large, but finite, set of data records. A *counting query* over $\mathcal{X}$ is given by a predicate $f : \mathcal{X} \to \{0,1\}$. Every predicate f splits $\mathcal{X}$ into $\mathcal{X}_{f,0}$ and $\mathcal{X}_{f,1}$ where, for $b = 0,1$, $\mathcal{X}_{f,b} = \{x \in \mathcal{X} | \ f(x) = b\}$. The predicate f is called *balanced* if $|\mathcal{X}_{f,0}| = |\mathcal{X}_{f,1}|$. Let $\mathcal{F}$ be a finite class of counting queries over $\mathcal{X}$. It is called *balanced* if every $f \in \mathcal{F}$ is balanced.

A *database* over $\mathcal{X}$ is a tuple $D \in \mathcal{X}^n$ for some $n \in \mathbb{N}$. n is called the *size* of D. Given a counting query f and a database $D = (d_1, \ldots, d_n) \in \mathcal{X}^n$, we may think of $\bar{f}(D) := \frac{1}{n}\sum_{i=1}^n f(d_i)$ as the correct answer.

Definition 1 ([6]). *Let $\mathcal{R}$ be a (possibly infinite) set. A random map $\mathcal{M} : \mathcal{X}^* \to \mathcal{R}$ (meaning that, for every $D \in \mathcal{X}^*$, $\mathcal{M}(D)$ is an $\mathcal{R}$-valued random variable) is said to provide* ε-differential privacy *if, for every $n \in \mathbb{N}$, for every pair $(D, D') \in \mathcal{X}^n \times \mathcal{X}^n$ of databases differing in one entry only, and for every measurable $S \subseteq \mathcal{R}$, the condition*

$$\Pr[\mathcal{M}(D) \in S] \le e^\varepsilon \Pr[\mathcal{M}(D') \in S]$$

is valid.

Random maps, sometimes called *database access mechanisms* in this context, are used for generating “noisy answers” to queries. On one hand, the amount of noise should be large enough for providing ε-differential privacy. On the other hand, the noise should not torpedo the usefulness of the answers. In this paper we focus on *non-interactive* database access mechanisms meaning that the same mechanism is used for a whole class $\mathcal{F}$ of counting queries. Useful non-interactive database access mechanisms are harder to design than their interactive counterparts[1] but have the advantage that the information $\mathcal{M}(D)$ can be computed once for the whole class $\mathcal{F}$ and be made public. However, for obvious reasons, the party that transforms D into $\mathcal{M}(D)$ must be trusted. In absence of a trusted party, one may need an even more restricted access mechanism, which is defined as follows:

Definition 2 ([6]). *A* randomized response scheme *is a database access mechanism $\mathcal{M}$ of the form*

$$\mathcal{M}(d_1, \ldots, d_n) = (Z(d_1), \ldots, Z(d_n)) \tag{3}$$

for some random map $Z : \mathcal{X} \to \mathcal{R}$.

Note that, in this case, $\mathcal{M}$ provides ε-differential privacy if and only if Z does (so that the random map Z is forced to blur almost all information in the individual data d_i). In applications, a randomized response scheme enables users to upload their data d_i in the noisy form $Z(d_i)$ on the database (without the need for a trusted party).

The following distribution has proved quite useful for providing ε-differential privacy:

Definition 3. *For any $\lambda > 0$, the* Laplace-distribution $\mathrm{Lap}(\lambda)$ *is the distribution over the reals that is given by the density function*

$$h(y) \propto \exp(-|y|/\lambda) \enspace .$$

*The d-*dimensional Laplace-distribution *is the product of d independent (1-dimensional) Laplace-distributions $\mathrm{Lap}(\lambda_i)$, $i = 1, \ldots, d$. It is denoted $\mathrm{Lap}(\vec{\lambda})$ for $\vec{\lambda} = (\lambda_1, \ldots, \lambda_d)$. If, for some $\lambda > 0$, $\lambda = \lambda_1 = \ldots = \lambda_d$, then we simply write $\mathrm{Lap}(\lambda)^d$ instead of $\mathrm{Lap}(\lambda, \ldots, \lambda)$.*

Adding Laplace-distributed noise to an $\mathbb{R}^d$-valued function f is effective if f has a low “sensitivity” in the sense of the following

[1] where the chosen random map may depend on the actual query, or even on all queries that happened in the past

Definition 4 ([6]). *The* sensitivity *of a function $f : \mathcal{X}^n \rightarrow \mathbb{R}^d$ is given by*

$$S(f) = \sup_{D,D'} \|f(D) - f(D')\|_1$$

where the supremum is taken over all $D, D' \in \mathcal{X}^n$ that differ in one entry only.

With these definitions, the following holds:

Lemma 2 ([6]). *Let $f : \mathcal{X}^n \rightarrow \mathbb{R}^d$ be a function of finite sensitivity, and let $Y \sim \mathrm{Lap}(S(f)/\varepsilon)^d$. Then, the random function F given by $F(D) = f(D)+Y$ provides ε-differential privacy.*

2.2 Prerequisites from Learning Theory

In learning theory, a class $\mathcal{F}$ of counting queries is usually called a *concept class* and the underlying universe $\mathcal{X}$ is called the *domain*. For our purpose, it will be convenient to deal with the *sign matrix* $A = (A[x, f]) \in \{-1, 1\}^{\mathcal{X}\times\mathcal{F}}$ such that $A[x, f] = 2f(x)-1$. As shown in [13], the (weak) learnability of $\mathcal{F}$ in the so-called SQ-model of learning [15] is closely related to linear arrangements $\mathcal{A}$ for the sign matrix A and the margins achieved by $\mathcal{A}$. In this section, we briefly remind the reader of these relations.

A *d-dimensional (homogeneous linear) arrangement for a sign matrix $A \in \mathbb{R}^{M\times N}$* is given by

$$\mathcal{A} = (u_1, \dots, u_M; v_1, \dots, v_N)$$

where u_i, v_j are vectors in $\mathbb{R}^d$ of unit Euclidean norm. With an arrangement $\mathcal{A}$ for A, we associate the *margin parameters* $\gamma_{i,j}(A|\mathcal{A}) = \langle u_i, v_j\rangle \cdot A_{i,j}$ and

$$\bar{\gamma}_j(A|\mathcal{A}) = \frac{1}{M}\sum_{i=1}^{M} \gamma_{i,j}(A|\mathcal{A}) \ ,$$
$$\bar{\gamma}_{min}(A|\mathcal{A}) = \min_{j=1,\dots,N} \bar{\gamma}_j(A|\mathcal{A}) \ .$$

Clearly, the value of a margin parameter lies in the interval $[-1, 1]$. The arrangement $\mathcal{A}$ for A is called *error-free* if none of the margin parameters $\gamma_{i,j}(A|\mathcal{A})$ is negative. The parameter $\bar{\gamma}_{min}(A|\mathcal{A})$ gives the row-average margin that is guaranteed by $\mathcal{A}$ for every column of A. Intuitively, we should think of an arrangement as being "good" if it induces "large margin" parameters. For this reason, we define

$$\bar{\gamma}_{min}(A) = \max_{\mathcal{A}} \bar{\gamma}_{min}(A|\mathcal{A}) \quad \text{and}$$
$$\bar{\gamma}^{ef}_{min}(A) = \max_{\mathcal{A}_{ef}} \bar{\gamma}_{min}(A|\mathcal{A}_{ef}) \ ,$$

where $\mathcal{A}_{ef}$ ranges over error-free arrangements for A only. Margin parameters exhibit nice relations to the so-called Forster bound. For our purpose the following variant [13] of the Forster bound [8] is needed:

$$\mathrm{FB}(A) = \max_q \mathrm{FB}_q(A) \quad \text{for} \quad \mathrm{FB}_q(A) = \frac{\sqrt{M}}{\|A \cdot \mathrm{diag}(q)^{1/2}\|_2} \tag{4}$$

Here, q ranges over all N-dimensional probability vectors. As shown in [3], the number of examples required to weakly learn $\mathcal{F}$ in the SQ-model under the uniform distribution on $\mathcal{X}$ is polynomially related to the SQ-dimension of $\mathcal{F}$.[2] We denote the latter as $\mathrm{SQdim}(A)$ (regarding the sign matrix A as our representation of $\mathcal{F}$). We are now prepared to describe the relations among the various parameters associated with a sign matrix.

Lemma 3 ([13]). *Let $A \in \{-1, 1\}^{M\times N}$ be a sign matrix. Then the following holds:*

1. *The parameters $\mathrm{SQdim}(A), \mathrm{FB}(A), \bar{\gamma}_{min}(A)^{-1}$ are pairwise polynomially related.*
2. *$\bar{\gamma}_{min}(A) = \max_{\mathcal{A}} \min_q \sum_{j=1}^{N} q_j \bar{\gamma}_j(A|\mathcal{A})$ where $\mathcal{A}$ ranges over all arrangements for A and q ranges over all probability vectors.*

The reciprocal value of a margin parameter, like $\bar{\gamma}_{min}(A)^{-1}$ in Lemma 3, is referred to as "margin complexity" ("small margin" = "high margin complexity").

2.3 Outline of the Main Results

Although we will not make it explicit very often, we consider the classes of counting queries as indexed by a "complexity parameter" k, i.e., $\mathcal{F} = \mathcal{F}_k$ (resp. $\mathcal{X} = \mathcal{X}_k$) is a member of a family $(\mathcal{F}_k)_{k\geq 1}$ (resp. $(\mathcal{X}_k)_{k\geq 1}$). As for Boolean predicates, k is typically the number of Boolean variables. Furthermore, we always assume that $\log|\mathcal{F}_k|$ and $\log|\mathcal{X}_k|$ are polynomially bounded in k. The sign matrix corresponding to $\mathcal{F}_k$ is denoted A_k.

Given $f \in \cup_{k\geq 1}\mathcal{F}_k$, $0 \leq \omega \leq 1$, and $n \geq 1$, we define an *(f, ω, n)-random database*, denoted by $X^n_{f,\omega}$, as the outcome of the following random experiment: draw ωn instances independently and uniformly at random from $\mathcal{X}_{f,1}$, and draw $(1 - \omega)n$ instances independently and uniformly at random from $\mathcal{X}_{f,0}$.

Informally, we consider a database access mechanism $\mathcal{M}$ "useful" for $\cup_{k\geq 1}\mathcal{F}_k$ if, for every counting query f, there exists a function Q_f such that (at least for random databases D) the "noisy answer" $Q_f(\mathcal{M}(D))$ is (in probability) a "good approximation" of the "correct answer" $\bar{f}(D)$. Note that $\bar{f}(D) = \omega$ if D is an (f, ω, n)-random database. Moreover, the required database size should be polynomially bounded in k (and in some other relevant variables). This is captured more formally in the following

Definition 5. *We say that a database access mechanism $\mathcal{M}$ is* strongly useful *for the family $(\mathcal{F}_k)_{k\geq 1}$ if there exists a function $n = n(k, \alpha, \beta)$ that is polynomially bounded in $k, 1/\alpha, 1/\beta$ and if there exists a family $(Q_f)_{f\in\cup_{k\geq 1}\mathcal{F}_k}$ of (possibly random) maps such that the following holds. For every $k \geq 1$, for every counting query $f \in \mathcal{F}_k$, for all $0 \leq \omega \leq 1$ and all $0 < \alpha, \beta \leq 1/2$, if $n \geq n(k, \alpha, \beta)$, $D = X^n_{f,\omega}$ and $\hat{\omega} = Q_f(\mathcal{M}(D))$, then the probability for $|\hat{\omega} - \omega| > \alpha$ is bounded by β. Here the probability is taken over the internal randomization in D, $\mathcal{M}$ and Q_f.*

We refer to $n(k, \alpha, \beta)$ as the *sample size required by $\mathcal{M}$*. We briefly note that the main difference between our notion of strong usefulness and the original notion of (α, β)-usefulness in [4] is that we deal with an (f, ω, n)-random database whereas, in [4], the criterion in terms of α, β has to hold for *all* choices of D.

The definition of *weak usefulness* is similar except that ω is set to either 0 or 1 and it suffices that, for each choice of $\omega, \beta, \varepsilon, f$, the probability for $|\hat{\omega} - \omega| < 1/2$ is at least $1 - \beta$ (so that the cases $\omega = 0$ and $\omega = 1$ can be distinguished with a high probability of success).

[2] In this paper, we will not need a formal definition of the SQ-model.

Definition 6. *We say that a database access mechanism $\mathcal{M}$ is* weakly useful *for the family $(\mathcal{F}_k)_{k\geq 1}$ if there exists a function $n = n(k,\beta)$ that is polynomially bounded in $k, 1/\beta$ and if there exists a family $(Q_f)_{f\in\cup_{k\geq 1}\mathcal{F}_k}$ of (possibly random) maps such that the following holds. For every $k \geq 1$, for every counting query $f \in \mathcal{F}_k$, for all $\omega \in \{0,1\}$ and all $0 < \beta \leq 1/2$, if $n \geq n(k,\beta)$, $D = X^n_{f,\omega}$ and $\hat{\omega} = Q_f(\mathcal{M}(D))$, then the probability for $|\hat{\omega} - \omega| \geq 1/2$ is bounded by β.*

The results in [6] imply that, for every $\varepsilon > 0$, there can be no randomized response scheme for the class of parity functions that provides ε-differential privacy and is weakly useful. The main results in this paper are as follows:

Theorem 1. *For every $\varepsilon > 0$, the following holds. If the margin complexity $\bar{\gamma}_{min}(A_k)^{-1}$ associated with a family $(\mathcal{F}_k)_{k\geq 1}$ of concept classes is a super-polynomial function in k, then $(\mathcal{F}_k)_{k\geq 1}$ has no randomized response scheme that provides ε-differential privacy and is weakly useful.*

The proof of this result is given in Sections 3 and 4. The proof of the next one will be given in Section 5.

Theorem 2. *For every $\varepsilon > 0$, there is a randomized response scheme $\mathcal{M}_\varepsilon$ for $(\mathcal{F}_k)_{k\geq 1}$ that provides ε-differential privacy and is strongly useful if at least one of the following conditions is valid:*

- *The classes $\mathcal{F}_k$ are balanced and $\bar{\gamma}_{min}(A_k)^{-1}$ is polynomially bounded in k.*
- *$\bar{\gamma}^{ef}_{min}(A_k)^{-1}$ is polynomially bounded in k.*

Moreover, there is a single polynomial[3] *in $k, 1/\alpha, 1/\beta, 1/\varepsilon$ that bounds from above the sample size required by $\mathcal{M}_\varepsilon$.*

It can be shown that Theorem 2 applies, for instance, to Boolean Monomials, Boolean Clauses or, more generally, to "Boolean Decision Lists with a bounded number of label-changes". It applies furthermore to "Axis-Parallel Hyper-Rectangles over a discrete domain". This is explained in more detail in Section 6.

Our definitions of weak and strong usefulness ignore efficiency issues. However, we will briefly indicate in Section 7 under which conditions the randomized response scheme from Theorem 2 can be implemented efficiently. This will specifically be the case for the aforementioned function classes.

2.4 Comparison to Prior Work

A first characterization of non-interactive private data analysis in the "central" model, where the sensitive database is managed by a trusted curator, is provided in [4]. Specifically, the authors show that, ignoring computational constraints, it is possible to privately release sanitized databases so as to provide (α,β)-usefulness for any concept class with polynomial VC-dimension.

Private learning in the absence of a trusted curator is addressed by a series of prior works (see for example [14, 5]). In particular, Kasiviswanathan et al. [14] prove that a concept class is learnable in the statistical query model if and only if it is learnable by what they call a local algorithm. More specifically, they show that there is an efficient mutual simulation between the respective oracles for these models: the SQ-oracle and the so-called LR-oracle. A non-adaptive local algorithm in the sense of [14] corresponds to an algorithm which has access to the data via a randomized response scheme in the sense of Definition 2. We would like to stress the following differences between the results in [14] and our main results:

- The SQ-learner (resp. the equivalent local algorithm) gets statistical information about a database consisting of data which are *labeled* according to an *unknown target function f*. In combination with a result from [3], this offers the possibility to design a non-adaptive SQ-learner (resp. a non-adaptive local algorithm) which weakly learns $(\mathcal{F}_k)_{k\geq 1}$ under the uniform distribution provided that the SQ-dimension of $\mathcal{F}_k$ grows polynomially with k only.
- In our setting, we have a *known query function f* in the role of the unknown target function, but the randomized response scheme is applied to *unlabeled data*. It should be noted that, even if the SQ-dimension of $\mathcal{F}_k$ is polynomially bounded, there is no direct way to transform a given non-adaptive SQ-learner into a successful randomized response scheme of the form (3) in our setting. The transformation from [14] cannot be used here because it assumes a labeled database. The main problem in our setting (where $D = (d_1,\ldots,d_n)$ consists of unlabeled data) is to find a *single random map Z* such that the tuple $\mathcal{M}(D) = (Z(d_1),\ldots,Z(d_n))$, despite of having a low sensitivity, contains enough information about the labels of *all query functions $f \in \mathcal{F}_k$*. We will see in Section 5 that using a linear arrangement for $\mathcal{F}_k$ is the key for solving this problem.

In [10], Gupta et al. show that if a database access mechanism can only access the database by means of the SQ-oracle (or, equivalently, by means of the LR-oracle), then the concept class can be released so as to provide ε-differential privacy and (α,β)-usefulness if and only if it is *agnostically SQ-learnable*. The main difference between these results and our contribution consists in the notion of usefulness that the randomized response scheme must achieve. The work of Gupta et al. requires the private mechanism to output values which are α-close to the true answers with probability at least $1-\beta$ for *all* possible choices of the input database D, while we are satisfied if this property holds for (f,ω,n)-random databases. In other words, we have replaced the worst-case analysis in [10] by a kind of average-case analysis. While the worst-case analysis establishes the very high barrier of "agnostic SQ-learnability", our main results show that the barrier in the average case, namely weak SQ-learnability under the uniform distribution, is much lower.

3. PROOF OF THEOREM 1 FOR BALANCED CLASSES

Here we present a proof under the additional assumption that the classes $\mathcal{F}_k$ are balanced. This proof will be obtained in a (more or less) straightforward manner from Theorem 3 below (being valid for *all* classes $\mathcal{F}$), which generalizes a similar theorem in [6] (being concerned with the class of parity functions only). It turns out that the proof of Theorem 3 is quite similar to the proof of the less general result in [6]. The main new technical contribution is located in the proof of Lemma 4 where we apply some algebraic manipulations

[3] as opposed to a family $(p_\varepsilon)_{\varepsilon>0}$ of polynomials in $k, 1/\alpha, 1/\beta$

that bring the Forster bound into play, which finally leads us to the more general result.

Let $\mathcal{F} = \{f_1, \ldots, f_N\}$ be a class of counting queries over the universe $\mathcal{X} = \{x_1, \ldots, x_M\}$. $A \in \{-1,1\}^{\mathcal{X}\times\mathcal{F}}$ denotes the corresponding sign matrix. In the sequel, we assume that $\mathcal{F}$ is balanced. The general case is discussed in Section 4.

Let $q = (q(f))_{f\in\mathcal{F}}$ denote a probability vector. Drawing f at random from $\mathcal{F}$ according to q is denoted by $f \sim \mathcal{F}^q$. Let X denote the random variable that is uniformly distributed on $\mathcal{X}$. For every $f \in \mathcal{F}$ and every $b = 0,1$, $X_{f,b}$ denotes the random variable that is uniformly distributed on $\mathcal{X}_{f,b}$. With each query function $f \in \mathcal{F}$ we associate the two quite diverse databases $X^n_{f,0}, X^n_{f,1} \in \mathcal{X}^n$, which (as explained in Section 2.3) are the (f,ω,n)-random databases for $\omega = 0,1$. An useful answer to a query instantiated by f should be close to b if the underlying (random) database is $X^n_{f,b}$. Suppose that the answer to the query is derived from an ε-differentially private randomized response scheme $\mathcal{M}$ for some $\varepsilon > 0$. The following result indicates that the usefulness of the answer is in this case severely limited (at least for query classes whose Forster bound is large):

THEOREM 3. *If $\mathcal{M}$ is an ε-differentially private randomized response scheme for the balanced class $\mathcal{F}$, then (given the above notations) the following holds with a probability of at least 7/8 (taken over $f \sim \mathcal{F}^q$):*

$$\mathrm{SD}\left(\mathcal{M}(X^n_{f,0}), \mathcal{M}(X^n_{f,1})\right) \le 4nU^{1/3} \quad \text{for} \quad U = \frac{(e^\varepsilon-1)^2}{\mathrm{FB}_q(A)^2} \ . \tag{5}$$

PROOF. Let $Z : \mathcal{X} \to \mathcal{R}$ be the random map such that (3) holds. The following observations are made in [6] already:

1. The ε-differential privacy of $\mathcal{M}$ implies that
$$\forall x, x' \in \mathcal{X}, \forall z \in \mathcal{R}: \ \Pr[Z(x) = z] \le e^\varepsilon \Pr[Z(x') = z] \ . \tag{6}$$
2. For each $f \in \mathcal{F}$, the following holds:
   (a) $\mathrm{SD}\left(\mathcal{M}(X^n_{f,0}), \mathcal{M}(X^n_{f,1})\right) \le n \cdot \mathrm{SD}\left(Z(X_{f,0}), Z(X_{f,1})\right)$.
   (b) $\mathrm{SD}\left(Z(X_{f,0}), Z(X)\right) = \mathrm{SD}\left(Z(X_{f,1}), Z(X)\right) = \frac{1}{2} \cdot \mathrm{SD}\left(Z(X_{f,0}), Z(X_{f,1})\right)$.

It therefore suffices to show that the probability (taken over $f \sim \mathcal{F}^q$) for

$$\mathrm{SD}\left(Z(X_{f,1}), Z(X)\right) \le \left(\frac{U}{\alpha}\right)^{1/3} \tag{7}$$

is at least $1-\alpha$ for every choice of $0 < \alpha < 1$. This sufficient condition actually holds as will become evident from Lemmas 4 and 5 below. Setting $\alpha = 1/8$, the theorem follows. □

The proof of Theorem 3 made use of two lemmas that we present now. For sake of brevity, we define

$$\begin{aligned} h(z|x) &:= \Pr[Z(x) = z] \ , \\ h_{f,b}(z) &:= \Pr[Z(X_{f,b}) = z] \ , \\ h(z) &= \Pr[Z(X) = z] \ . \end{aligned} \tag{8}$$

Let $B \in_R \{0,1\}$ denote the Bernoulli distribution with parameter 1/2 such that $b \sim B$ is a perfect random bit.

LEMMA 4. *Suppose that Z is a random map satisfying condition (6). Then, for all z in the range of Z and for U as specified in (5), the following holds:*

$$\begin{aligned} \mathbb{E}_{f\sim\mathcal{F}^q, b\sim B}[h_{f,b}(z)] &= h(z) \quad \text{and} \\ \mathrm{Var}_{f\sim\mathcal{F}^q, b\sim B}[h_{f,b}(z)] &\le Uh(z)^2 \end{aligned} \tag{9}$$

PROOF. Let $h_{min}(z) = \min_{x\in\mathcal{X}} h(z|x)$, $\bar{h}(z|x) = h(z|x) - h_{min}(z)$, $\bar{h}_{f,b}(z) = h_{f,b}(z) - h_{min}(z)$, and $\bar{h}(z) = h(z) - h_{min}(z)$. Let furthermore $\vec{h(z)}$ be the vector $(\bar{h}(z|x))_{x\in\mathcal{X}}$. The following observations, partially made in [6] already, hold for every z in the range of Z:

$$\begin{aligned} \forall f \in \mathcal{F}: \mathbb{E}_{b\sim B}[h_{f,b}(z)] &= \frac{1}{2}(h_{f,0}(z) + h_{f,1}(z)) \\ &= h(z) && (10) \\ \mathbb{E}_{f\sim\mathcal{F}^q, b\sim B}[h_{f,b}(z)] &= h(z) && (11) \\ \mathrm{Var}_{f\sim\mathcal{F}^q, b\sim B}[h_{f,b}(z)] &= \mathrm{Var}_{f\sim\mathcal{F}^q, b\sim B}\left[\bar{h}_{f,b}(z)\right] && (12) \\ h(z|x) &\le e^\varepsilon h_{min}(z) && (13) \\ \|\vec{h(z)}\|_2^2 &\le M(e^\varepsilon-1)^2 h(z)^2 && (14) \end{aligned}$$

(10) holds because the functions $f \in \mathcal{F}$ are assumed as balanced. (11) is an immediate consequence of (10). (12) is obvious. (13) holds because Z satisfies (6). (14) follows from (13) and the following calculation:

$$\begin{aligned} \|\vec{h(z)}\|_2^2 &= \sum_x \bar{h}(z|x)^2 = \sum_x (h(z|x) - h_{min}(z))^2 \\ &\overset{(13)}{\le} M(e^\varepsilon-1)^2 h_{min}(z)^2 \le M(e^\varepsilon-1)^2 h(z)^2 \end{aligned}$$

For sake of brevity, we set $A_q = A \cdot \mathrm{diag}(q^{1/2})$ and proceed as follows:

$$\begin{aligned} \mathrm{Var}_{f\sim\mathcal{F}^q, b\sim B}[h_{f,b}(z)] &\overset{(12)}{=} \mathrm{Var}_{f\sim\mathcal{F}^q, b\sim B}\left[\bar{h}_{f,b}(z)\right] \\ &\overset{(11)}{=} \mathbb{E}_{f\sim\mathcal{F}^q, b\sim B}\left[\left(\bar{h}_{f,b}(z) - \bar{h}(z)\right)^2\right] \\ &= \mathbb{E}_{f\sim\mathcal{F}^q, b\sim B}\left[\left(\frac{2}{M}\sum_x \bar{h}(z|x)\mathbf{1}_{\mathcal{X}_{f,b}}(x) - \frac{1}{M}\sum_x \bar{h}(z|x)\right)^2\right] \\ &= \mathbb{E}_{f\sim\mathcal{F}^q, b\sim B}\left[\left(\frac{1}{M}\sum_x \bar{h}(z|x)(2\cdot\mathbf{1}_{\mathcal{X}_{f,b}}(x) - 1)\right)^2\right] \\ &= \mathbb{E}_{f\sim\mathcal{F}^q}\left[\left(\frac{1}{M}\sum_x \bar{h}(z|x)A[x,f]\right)^2\right] \\ &= \frac{1}{M^2}\sum_f q(f)\left(\sum_x \bar{h}(z|x)A[x,f]\right)^2 \\ &= \frac{1}{M^2}\sum_{x,x'} \bar{h}(z|x)\bar{h}(z|x')\sum_f q(f)A[x,f]A[x',f] \\ &= \frac{1}{M^2}\vec{h(z)}^\top A_q A_q^\top \vec{h(z)} \\ &\overset{(2)}{\le} \frac{1}{M^2}\|\vec{h(z)}\|_2^2 \|A_q\|_2^2 \overset{(14),(4)}{\le} \left(\frac{(e^\varepsilon-1)h(z)}{\mathrm{FB}_q(A)}\right)^2 \end{aligned}$$

The third equation in this calculation results from expanding the definitions of $\bar{h}_{f,b}(z)$ and $\bar{h}(z)$ and from making use of the fact that f is balanced. The fifth equation holds because $2\cdot\mathbf{1}_{\mathcal{X}_{f,b}}(x) - 1 = (2b-1)\cdot A[x,f]$. Since the expression containing the factor $2b-1 \in \{\pm 1\}$ is squared, the parameter

b vanishes at this stage of the computation. The remaining steps in the calculation are rather obvious. □

The following result (resp. a slightly less general version of it) is shown in [6]. For sake of completeness, the proof is given in Section A.

Lemma 5. *Let $h(z|x)$, $h_{f,b}(z|x)$ and $h(z)$ be given by (8), and let $Z : \mathcal{X} \rightarrow \mathcal{R}$ be a random map such that (9) holds for every z in the range of Z. Let $0 < \alpha < 1$. Then the probability for (7), taken over $f \sim \mathcal{F}^q$, is at least $1 - \alpha$.*

We are now ready to prove Theorem 1 under the assumption that the classes $\mathcal{F}_k$ are balanced. Let $\mathcal{M}$ be a randomized response scheme for $(\mathcal{F}_k)_{k\geq 1}$ that provides ε-differential privacy for some $\varepsilon > 0$. We will show that assuming $\mathcal{M}$ to be weakly useful *and* $\bar{\gamma}_{min}(A_k)^{-1}$ to be super-polynomial in k leads to a contradiction.
Suppose first that $\mathcal{M}$ is weakly useful. Let $\beta = 1/3$. Then, for all k and all $f \in \mathcal{F}_k$, the criterion for weak usefulness implies that the distributions $X^n_{f,0}$ and $X^n_{f,1}$ can be distinguished with a probability of at least 2/3 of success. Thus, there will be the same small average error rate of at most 1/3, when we q-average over all f from $\mathcal{F}_k$ regardless of how the distribution q on $\mathcal{F}_k$ is chosen.
Suppose second that $\bar{\gamma}_{min}(A_k)^{-1}$ is a super-polynomial function in k. Then, according to Lemma 3, $\mathrm{FB}(A_k)$ (as defined in (4)) is a super-polynomial function in k too. Thus, for every k, there must exist a probability vector $q = q_k$ such that $\mathrm{FB}_{q_k}(A_k)$ is a super-polynomial function in k. According to Theorem 3 (applied to $\mathcal{F} = \mathcal{F}_k$), inequality (5) is valid with a probability (taken over $f \sim \mathcal{F}^q$) of at least 7/8. Since the database size n may grow polynomially in k only, the statistical difference between $\mathcal{M}(X_{f,0})$ and $\mathcal{M}(X_{f,1})$ will be negligible (as k goes to infinity) for a q-fraction 7/8 (or more) of all $f \in \mathcal{F}_k$. We apply Lemma 1 and conclude that, for the "bad functions" $f \in \mathcal{F}_k$, the Bayes error for distinguishing between $\mathcal{M}(X_{f,0})$ and $\mathcal{M}(X_{f,1})$ will therefore be arbitrarily close to 1/2 (as k goes to infinity). Thus the Bayes error, q-averaged over all $f \in \mathcal{F}_k$ cannot be significantly smaller than $(7/8)\cdot(1/2) + (1/8)\cdot 0 = 7/16$.
Since $7/16 > 1/3$, we arrived at a contradiction.

4. PROOF OF THEOREM 1

In Section 3, we presented a proof of Theorem 1 under the additional assumption that the classes $\mathcal{F}_k$ are balanced. The proof was obtained in a (more or less) straightforward manner from Theorem 3 which presents an upper bound on $\mathrm{SD}(\mathcal{M}(X^n_{f,0}), \mathcal{M}(X^n_{f,1}))$. The main point in the proof was that this upper bound is asymptotically smaller than $1/P(k)$ for any polynomial P provided that the margin complexity $\bar{\gamma}_{min}(A_k)^{-1}$ associated with $(\mathcal{F}_k)_{k\geq 1}$ is super-polynomial in k. The main reasons why Theorem 1 holds even for unbalanced classes are as follows:

- Theorem 3 can be generalized to "almost balanced" classes. See Theorem 4 below for the formal statement.

- If the margin complexity $\bar{\gamma}_{min}(A_k)^{-1}$ associated with $(\mathcal{F}_k)_{k\geq 1}$ is super-polynomial in k, then there are almost balanced sub-classes of $\mathcal{F}_k$ whose margin complexity is still super-polynomial in k. See Theorem 5 below for the formal statement.

We define the *imbalance* of a function $f : \mathcal{X} \rightarrow \{0,1\}$ as $\left||\mathcal{X}_{f,1}| - |\mathcal{X}_{f,0}|\right|/|\mathcal{X}|$. We say that $\mathcal{F}$ has an *imbalance* of a most Δ if, for each $f \in \mathcal{F}$, the imbalance of f is at most Δ. Let $\mathcal{X}' \supseteq \mathcal{X}$ be an extended universe such that $|\mathcal{X}'| = |\mathcal{X}| + \Delta|\mathcal{X}|$. Clearly, the functions f of a class with an imbalance of at most Δ can be extended to the larger domain $\mathcal{X}'$ so as to become (completely) balanced. This balanced extension of f is denoted f' so that $\mathcal{F}' = \{f' : f \in \mathcal{F}\}$ is a balanced extension of $\mathcal{F}$. With these notations, the following holds:

Lemma 6. *For all $f \in \mathcal{F}$, the following holds:*

$$\mathrm{SD}(X_{f',0}, X_{f,0}) + \mathrm{SD}(X_{f',1}, X_{f,1}) \leq 2\Delta \ .$$

The (straightforward) proof is omitted.

Lemma 7. *Suppose that $\mathcal{F}$ has an imbalance of at most Δ. Let $\mathcal{F}'$ be its balanced extension, and let A and A' be the corresponding sign matrices, respectively. Let q be a probability vector for the functions in $\mathcal{F}$. With these notations, the following holds:*

$$\frac{1}{\mathrm{FB}_q(A')^2} \leq \frac{1}{\mathrm{FB}_q(A)^2} + \frac{2\sqrt{\Delta}}{\mathrm{FB}_q(A)} + \Delta \ .$$

Proof. For sake of brevity, we set $A_q = A \cdot \mathrm{diag}(q)^{1/2}$. We may think of A' as having the sign matrix A (with $|\mathcal{X}|$ rows) in its upper block and another sign matrix, say E, with $\Delta|\mathcal{X}|$ rows in its lower block. With these notations, the spectral norm of A'_q can be bounded from above as follows:

$$\|A'_q\|_2 \leq \|A_q\|_2 + \|E_q\|_2 \overset{(2)}{\leq} \|A_q\|_2 + ||E_q||_2 \overset{(1)}{\leq} \|A_q\|_2 + \sqrt{\Delta M}$$

Thus, we obtain

$$\begin{aligned} \frac{1}{\mathrm{FB}_q(A')^2} &\overset{(4)}{=} \frac{\|A'_q\|_2^2}{(1+\Delta)M} \leq \frac{(\|A_q\|_2 + \sqrt{\Delta M})^2}{(1+\Delta)M} \\ &\leq \frac{\|A_q\|_2^2}{M} + \frac{2\sqrt{\Delta M}\|A_q\|_2}{M} + \frac{\Delta M}{M} \\ &\overset{(4)}{=} \frac{1}{\mathrm{FB}_q(A)^2} + \frac{2\sqrt{\Delta}}{\mathrm{FB}_q(A)} + \Delta \ , \end{aligned}$$

which concludes the proof. □

We now arrive at the following extension of Theorem 3:

Theorem 4. *Suppose that $\mathcal{F}$ has an imbalance of at most Δ. Let $\mathcal{F}'$ be its balanced extension, and let A and A' be the corresponding sign matrices, respectively. Let q be a probability vector for the functions in $\mathcal{F}$, and let $\varepsilon > 0$. If $\mathcal{M}$ is an ε-differentially private randomized response scheme, then the following holds with a probability of at least 7/8 (taken over $f \sim \mathcal{F}^q$):*

$$\mathrm{SD}\left(\mathcal{M}(X^n_{f,0}), \mathcal{M}(X^n_{f,1})\right) \leq 2\Delta n + 4n(e^{\varepsilon} - 1)^2 \cdot B \ ,$$

where

$$B = \frac{1}{\mathrm{FB}_q(A)^2} + \frac{2\sqrt{\Delta}}{\mathrm{FB}_q(A)} + \Delta \ .$$

Proof. According to the triangle inequality,

$$\mathrm{SD}(\mathcal{M}(X^n_{f,0}), \mathcal{M}(X^n_{f,1}))$$

is upper-bounded by

$$\sum_{b=0,1} \mathrm{SD}\left(\mathcal{M}(X^n_{f,b}), \mathcal{M}(X^n_{f',b})\right) + \mathrm{SD}\left(\mathcal{M}(X^n_{f',0}), \mathcal{M}(X^n_{f',1})\right) \ .$$

The first sum is upper-bounded by

$$\sum_{b=0,1} \mathrm{SD}\left(X^n_{f,b}, X^n_{f',b}\right) \le n \cdot \sum_{b=0,1} \mathrm{SD}\left(X_{f,b}, X_{f',b}\right)$$

and therefore, by virtue of Lemma 6, by $2\Delta n$. According to Theorem 3, the following holds with a probability of at least 7/8 (taken over $f \sim \mathcal{F}^q$):

$$\mathrm{SD}\left(\mathcal{M}(X^n_{f',0}), \mathcal{M}(X^n_{f',1})\right) \le \frac{4n(e^\varepsilon - 1)^2}{\mathrm{FB}_q(A')^2}$$

Plugging in the upper bound on $1/\mathrm{FB}_q(A')$ from Lemma 7 and putting all pieces together, the theorem follows. □

Here comes the final piece of the puzzle in our proof of Theorem 1:

Theorem 5. *If the margin complexity $\bar{\gamma}_{min}(A_k)^{-1}$ associated with a family $(\mathcal{F}_k)_{k\ge1}$ of concept classes is a super-polynomial function in k, then, for each polynomial $P(k) \ge 1$, there exists a family of sub-classes $\mathcal{F}^P_k \subseteq \mathcal{F}_k$ such that the following holds:*

1. *For each $k \ge 1$, the imbalance of $\mathcal{F}^P_k$ is bounded by $1/P(k)$.*
2. *The margin complexity $\bar{\gamma}_{min}(A^P_k)^{-1}$ associated with the family $(\mathcal{F}^P_k)_{k\ge1}$ is a super-polynomial function in k.*

Proof. We simply choose $\mathcal{F}^P_k$ as the set of all $f \in \mathcal{F}_k$ whose imbalance is bounded by $1/P(k)$. Let k be arbitrary but fixed. Let $\mathcal{F} = \mathcal{F}_k$ and $A = A_k$. Each, say d-dimensional, arrangement $\mathcal{A}$ for the sign matrix A can be extended as follows:

- Add an extra dimension $d + 1$.
- If $f \in \mathcal{F}^P$, then the coordinate $d + 1$ of the vector v_f is set to 0.
- If $f \in \mathcal{F}\backslash\mathcal{F}^P$ is biased towards positive (resp. negative) examples, then the coordinate $d + 1$ of v_f gets value 1 (resp. -1), and the remaining coordinates are set to 0.
- The coordinate $d + 1$ of each vector u_x is set to 1.
- For each $x \in \mathcal{X}$, the vector u_x is normalized so as to have unit Euclidean norm (by applying the scaling factor $1/\sqrt{2}$).

The effect is as follows:

- For each function $f \in \mathcal{F} \setminus \mathcal{F}^P$, the average margin achieved by the extension of $\mathcal{A}$ is at least $(\sqrt{2}P(k))^{-1}$.
- For each function $f \in \mathcal{F}^P$, the average margin achieved by the extension of $\mathcal{A}$ coincides with $1/\sqrt{2}$ times the average margin achieved by the original arrangement $\mathcal{A}$.

Since $\bar{\gamma}_{min}(A_k)^{-1}$ is super-polynomial in k but the functions from $\mathcal{F} \setminus \mathcal{F}^P$ cannot be blamed for it, it follows that $\bar{\gamma}_{min}(A^P_k)^{-1}$ is super-polynomial in k. □

The proof of Theorem 4 can now be completed in a similar fashion as it was done at the end of Section 3 for the special case of balanced classes. We omit the details.

5. PROOF OF THEOREM 2

The analysis of the randomized response scheme that we are going to design requires Hoeffding's inequality [11, 18], and, in addition, the following tail bound:

Theorem 6. *For $i = 1, \dots, n$ and $j = 1, \dots, d$, let $Y_{i,j} \sim \mathrm{Lap}(\lambda)$ be i.i.d. random variables. For each $i = 1, \dots, n$, let $v_i = (v_{i,j})_{j=1,\dots,d}$ be a vector of unit norm, i.e., $\|v_i\|_2 = 1$. Then, for each $0 < \delta \le \sqrt{8}\lambda n$, the following holds:*

$$\Pr\left[\sum_{i=1}^n \sum_{j=1}^d v_{i,j} Y_{i,j} \ge \delta\right] \le \exp\left(-\frac{\delta^2}{8n\lambda^2}\right)$$

Since we were not able to find a proper reference, we will give the proof of Theorem 6 (plus some additional remarks) in Section B. Note that the assumption $\delta \le \sqrt{8}\lambda n$ in Theorem 6 is not very restrictive because, for $\delta \ge \sqrt{8}\lambda n$, we get

$$\Pr\left[\sum_{i=1}^n \sum_{j=1}^d v_{i,j} Y_{i,j} \ge \delta\right] \le \Pr\left[\sum_{i=1}^n \sum_{j=1}^d v_{i,j} Y_{i,j} \ge \sqrt{8}\lambda n\right] \le e^{-n} \ .$$

We are now ready for the central part of this section. As usual, let $\mathcal{F}$ denote a class of counting queries over the universe $\mathcal{X}$. $A \in \{-1, 1\}^{\mathcal{X}\times\mathcal{F}}$ denotes the corresponding sign matrix. Let $\mathcal{A} = ((u_x)_{x\in\mathcal{X}}, (v_f)_{f\in\mathcal{F}})$ be a d-dimensional arrangement for A. We aim at designing a randomized response scheme that is strongly useful and provides ε-differential privacy. To this end, we define the random map $Z : \mathcal{X} \to \mathbb{R}^d$ according to

$$Z(x) = u_x + Y \quad \text{for} \quad Y \sim \mathrm{Lap}\left(\frac{2\sqrt{d}}{\varepsilon}\right)^d \ . \tag{15}$$

The random vectors Y are chosen independently for different choices of x.

Lemma 8. *Let $\varepsilon > 0$. Choose the random map Z according to (15) and $\mathcal{M}$ according to (3). Then, $\mathcal{M}$ provides ε-differential privacy.*

Proof. Let $D' \in \mathcal{X}^n$ be a second database differing from D in one entry only, say in entry $d'_k \ne d_k$. Consider the map $h(D) = (u_{d_1}, \dots, u_{d_n}) \in (\mathbb{R}^d)^n$. Then

$$\|h(D) - h(D')\|_1 = \|u_{d_k} - u_{d'_k}\|_1 \le \sqrt{d}\|u_{d_k} - u_{d'_k}\|_2 \le 2\sqrt{d} \ .$$

It follows that the sensitivity of h is bounded by $2\sqrt{d}$. According to Lemma 2 (applied with nd in place of d), the mechanism $\mathcal{M}$ provides ε-differentially privacy. □

For sake of simplicity, we will use the abbreviations $\bar{\gamma}(f) = \bar{\gamma}_f(A|\mathcal{A})$ and $\bar{\gamma}_{min} = \bar{\gamma}_{min}(A|\mathcal{A})$ in what follows. As in the definition of strong usefulness, $X^n_{f,\omega} \in \mathcal{X}^n$ denotes an (f, ω, n)-random database. Let $f \in \mathcal{F}$ be a counting query and $v = v_f$ its representation in the arrangement $\mathcal{A}$. The following margin parameters, associated with f (resp. v) and $b = 0, 1$, will play a central role:

$$\bar{\gamma}_b(f) = \frac{1}{|\mathcal{X}_{f,b}|} \sum_{x\in\mathcal{X}_{f,b}} (2b - 1)\langle u_x, v\rangle \ , \tag{16}$$

$$\bar{\gamma}(f) = \frac{1}{2}(\bar{\gamma}_0(f) + \bar{\gamma}_1(f)) \tag{17}$$

If the random database $D = X^n_{f,\omega}$ contains the instances $X_1, \ldots, X_n$, then its noisy version $\hat{D} = \mathcal{M}(D)$ has the following form:

$$\hat{D} = (\hat{X}_1, \ldots, \hat{X}_n) \ \text{ for } \ \hat{X}_i = Z(X_i) \overset{(15)}{=} u_{X_i} + Y_i$$
$$\text{and } Y_i \sim \text{Lap}\left(\frac{2\sqrt{d}}{\varepsilon}\right)^d .$$

Consider the random variable $S = S_1 + S_2$ for

$$S_1 = \frac{1}{n}\sum_{i=1}^{n} \langle u_{X_i}, v\rangle \ \text{ and } \ S_2 = \frac{1}{n}\sum_{i=1}^{n}\langle Y_i, v\rangle \ . \tag{18}$$

Now, we obtain

$$\mathbb{E}[S] = \mathbb{E}[S_1] = \omega\bar{\gamma}_1(f) - (1-\omega)\bar{\gamma}_0(f)$$
$$= \omega(\bar{\gamma}_0(f) + \bar{\gamma}_1(f)) - \bar{\gamma}_0(f) \ .$$

Solving for ω yields

$$\omega = \frac{\mathbb{E}[S] + \bar{\gamma}_0(f)}{\bar{\gamma}_0(f) + \bar{\gamma}_1(f)} \ . \tag{19}$$

Treating S as a (hopefully good) approximation of $\mathbb{E}[S]$, this motivates the following definition of $\hat{\omega}$:

1. Given $\hat{D}$, compute $S = S_1 + S_2 = \frac{1}{n}\sum_{i=1}^{n}\left\langle \hat{X}_i, v\right\rangle$.
2. Compute $\hat{\omega} = Q_f(\hat{D})$ according to the right hand-side of (19) except that S is substituted for (the unknown) $\mathbb{E}[S]$.

View now $\mathcal{F} = \mathcal{F}_k$ as a member of the parametrized class $(\mathcal{F}_k)_{k\geq 1}$. Remember that we assume $\log|\mathcal{X}_k|$ and $\log|\mathcal{F}_k|$ to be polynomially bounded in k. Note that $\bar{\gamma}(f) = \tilde{\gamma}(f)$ if f is balanced, and $\bar{\gamma}(f) \leq 2\tilde{\gamma}(f)$ if the arrangement $\mathcal{A}$ is error-free. Therefore, Theorem 2 will be an immediate consequence of the following two results.

THEOREM 7. *Let $\varepsilon > 0$. The randomized response scheme, given by Z and $(Q_f)_{f\in\mathcal{F}}$ as defined above, provides ε-differential privacy. Moreover, it is strongly useful provided that the size $n = n(d, \alpha, \beta, \varepsilon, \tilde{\gamma}_{min})$ of the (random) database satisfies*

$$n \geq \frac{32\cdot d}{\alpha^2\varepsilon^2\tilde{\gamma}^2_{min}}\ln\left(\frac{2}{\beta}\right) \tag{20}$$

where $\tilde{\gamma}_{min} = \min_{f\in\mathcal{F}}\tilde{\gamma}(f)$.

PROOF. ε-differential privacy is granted by Lemma 8. We have to show that, with probability at least $1-\beta$, $|\hat{\omega} - \omega| \leq \alpha$ where $\hat{\omega}$ is calculated according to the right hand-side of (19) except that S is substituted for $\mathbb{E}[S]$. An inspection of (19) shows that the condition $|\hat{\omega} - \omega| \leq \alpha$ is equivalent to $|S - \mathbb{E}[S]| \leq \alpha(\bar{\gamma}_0(f) + \bar{\gamma}_1(f))$. Since $S = S_1 + S_2$ and $2\tilde{\gamma}_{min} \leq 2\tilde{\gamma}(f) = \bar{\gamma}_0(f) + \bar{\gamma}_1(f)$, the latter condition is guaranteed if the following holds:

$$|S_1 - \mathbb{E}[S_1]| \leq \alpha\tilde{\gamma}_{min} \ \text{ and } |S_2 - \mathbb{E}[S_2]| \leq \alpha\tilde{\gamma}_{min} \tag{21}$$

According to (18), S_1 is the average over the independent random variables $\langle u_{X_i}, v\rangle \in [-1,1]$. An application of Hoeffding's inequality shows that the probability for the first condition in (21) to be violated is bounded by

$$2\exp(-\alpha^2\tilde{\gamma}^2_{min}n/2).$$

According to Theorem 6, the probability for the second condition in (21) to be violated is bounded by

$$2\exp(-\alpha^2\tilde{\gamma}^2_{min}\varepsilon^2 n/(32d)),$$

a bound which is more restrictive than the preceding one.[4] Setting

$$2\exp\left(\frac{-\alpha^2\tilde{\gamma}^2_{min}\varepsilon^2 n}{32d}\right) \leq \beta$$

and solving for n, we see that the database size n, as specified in (20), is sufficiently large. □

It is assumed in Theorem 2 that there are arrangements $\mathcal{A}_k$ for the classes $\mathcal{F}_k$ such that the margin complexity

$$\bar{\gamma}_{min}(A_k|\mathcal{A}_k)^{-1}$$

is polynomially bounded in k. As already noted above, $\bar{\gamma}_{min}$ and $\tilde{\gamma}_{min}$ are equal up to a factor of at most 2 under the assumption (made in Theorem 2) that the classes $\mathcal{F}_k$ are balanced or that the arrangements $\mathcal{A}_k$ are error-free. An inspection of (20) shows that the only remaining parameter with a possibly super-polynomial growth in k (or in the other relevant parameters) is the dimension $d = d(\mathcal{F}_k)$ of the arrangement. However, d can be assumed as small because of the following result:

THEOREM 8. *Given a (possibly high-dimensional) arrangement $\mathcal{A}$ for $\mathcal{F}$ such that $\gamma := \tilde{\gamma}_{min}(A|\mathcal{A}) > 0$, there exists a d-dimensional arrangement $\mathcal{A}'$ for $\mathcal{F}$ such that $\tilde{\gamma}_{min}(A|\mathcal{A}') \geq \gamma/8$ provided that*

$$d \geq 8\ln\left(\frac{4(M+N)}{\beta}\right) + \frac{128}{\gamma^2}\ln\left(\frac{8N}{\beta}\right) \tag{22}$$

The proof, which is based on random projection techniques [12, 1, 2], is found in Section C.

6. ARRANGEMENTS WITH A NON-NEGLIGIBLE MINIMUM MARGIN

In this section, we start with a (more or less trivial) observation, Lemma 9 below, which is applied afterward to a couple of Boolean concept classes.

Let $A \in \{-1,1\}^{M\times N}$ be a sign matrix and let $\mathcal{A}$ be an arrangement for A. The margin parameter $\bar{\gamma}_{min}(A|\mathcal{A})$ that is used in Theorems 1 and 2 is clearly bounded from below by

$$\gamma_{min}(A|\mathcal{A}) = \min\{\gamma_{i,j}(A|\mathcal{A}) : i = 1,\ldots,M \ , \ j = 1,\ldots,N\} \ ,$$

which is the minimum among all MN individual margin parameters.

Let $(\mathcal{F}_k)_{k\geq 1}$ be a family of concept classes, and let $A_k \in \{-1,1\}^{\mathcal{X}_k\times\mathcal{F}_k}$ be the sign matrix associated with $\mathcal{F}_k$. We say that $(\mathcal{F}_k)_{k\geq 1}$ has a *non-negligible minimum margin* if there exists a polynomial $P(k) \in \mathbb{N}$, and, for all $k \geq 1$, an arrangement $\mathcal{A}_k$ such that $\gamma_{min}(A_k|\mathcal{A}_k) \geq 1/P(k)$. This clearly implies that each arrangement $\mathcal{A}_k$ for A_k is error-free and that $\gamma_{min}(A_k|\mathcal{A}_k) \leq \bar{\gamma}^{ef}_{min}(A_k|\mathcal{A}_k)$. It follows that the conclusion of Theorem 2 holds, in particular, for all classes with a non-negligible minimum margin.

[4] Note that the application of Theorem 6 is justified because the assumption made there, namely $\delta \leq \sqrt{8}\lambda n$, is granted in our application where $\delta = \alpha\tilde{\gamma}_{min} \leq \alpha \leq 1$ and $\lambda = 2\sqrt{d}/\varepsilon \geq 2$.

For each $k \geq 1$, let $\mathcal{F}_k$ be a class of concepts over the Boolean domain $\mathcal{X}_k = \{0,1\}^k$ or $\mathcal{X}_k = \{-1,1\}^k$. We say that $(\mathcal{F}_k)_{k \geq 1}$ is *linearly separable with polynomially bounded weights* if there exists a polynomial $P(k) \in \mathbb{N}$, and, for all $k \geq 1$ and all $f \in \mathcal{F}_k$, a weight vector

$$v_f \in \{-P(k), \ldots, 0, \ldots, P(k)\}^k$$

together with a threshold $t \in \mathbb{R}$ such that, for all $x \in \mathcal{X}_k$,

$$f(x) = 1 \Leftrightarrow \langle v_f, x \rangle \geq t \ . \tag{23}$$

Assume now that this is the case. Clearly, the threshold t can w.l.o.g. be chosen from the interval $[-kP(k), kP(k)]$. It follows that $(\mathcal{F}_k)_{k \geq 1}$ has a non-negligible minimum margin. We could, for example, represent x and f by the $(k+1)$-dimensional vectors

$$\begin{aligned} x' &= \left(\sqrt{P(k)} \cdot x, \sqrt{kP(k)}\right) \quad \text{and} \\ v'_f &= \left(\frac{1}{\sqrt{P(k)}} \cdot v_f, -\frac{t}{\sqrt{kP(k)}}\right) \ , \end{aligned} \tag{24}$$

respectively, so that $\langle v'_f, x' \rangle = \langle v_f, x \rangle - t$. Moreover, t can be chosen in the middle between two consecutive non-negative integers. Denote the arrangement obtained by this setting as $\mathcal{A}_k$. Note that $\|v'_f\|_2, \|x'\|_2 \leq \sqrt{2kP(k)}$. After normalizing all vectors to unit Euclidean norm, we obtain $\gamma_{min}(A_k|\mathcal{A}_k) \geq 1/(4kP(k))$. We arrive at the following result:

Lemma 9. *Let $P(k)$ be a polynomial. If $(\mathcal{F}_k)_{k \geq 1}$ is a family of Boolean concept classes that is linearly separable with integer weights of absolute value at most $P(k)$, then there exist arrangements $(\mathcal{A}_k)_{k \geq 1}$ such that $\gamma_{min}(A_k|\mathcal{A}_k) \geq 1/(4kP(k))$.*

We now discuss some applications.

Boolean Monomials.

A Boolean monomial is linearly separable with weights from $\{-1, 0, 1\}$ (in the obvious fashion). Thus, Lemma 9 applies with $P(k) = 1$ for all $k \geq 1$. We can conclude that, for monomials over k Boolean variables, the minimum margin is at least $1/(4k)$, resp. at least $1/(2k)$ as it had been observed in [9].

Discrete Axis-Parallel Hyper-Rectangles.

Let $J(k) \in \mathbb{N}$ be polynomially bounded in k, and let $\mathcal{X}_k = \{1, \ldots, J(k)\}^k$. An axis-parallel hyper-rectangle over $\mathcal{X}_k$, also called "k-dimensional box", is a subset of the form

$$\{a_1, \ldots, b_1\} \times \ldots \times \{a_k, \ldots, b_k\}$$

where $a_i, b_i \in \{1, \ldots, J(k)\}$. Using the predicates $x_i \geq a$ and $x_i \leq b$ for $a, b \in \{1, \ldots, J(k)\}$, it is easy to cast a k-dimensional box as a monomial over $2kJ(k)$ Boolean variables. It follows that, for boxes of this form, the minimum margin is at least $1/(4k^2 J(k))$.

Boolean Decision Lists.

For a Boolean variable z, let $z^1 = z$ and $z^0 = \bar{z}$. We consider decision lists[5] over the Boolean variables $z_1, \ldots, z_k$ of the form

$$[(z_{i(1)}^{\varepsilon_1}, b_1), \ldots, (z_{i(s)}^{\varepsilon_s}, b_s), b_{s+1}] \tag{25}$$

where $1 \leq s \leq k$, $i(1), \ldots, i(s) \in \{1, \ldots, k\}$, and $\varepsilon_1, \ldots, \varepsilon_s$, $b_1, \ldots, b_s \in \{0,1\}$. A list of this form represents a Boolean function that assigns a bit to each vector $x \in \mathcal{X}_k = \{0,1\}^k$ as follows:

- If x does not satisfy any literal $z_{i(j)}^{\varepsilon_j}$ in the list, then return the bit b_{s+1}.
- Otherwise, let j be smallest index such that x satisfies the literal $z_{i(j)}^{\varepsilon_j}$. Return b_j.

It is well known that Boolean decision lists with a bounded number of label changes (from 0 to 1 or vice versa) within the sequence $b_1, \ldots, b_s, b_{s+1}$ are linearly separable with polynomially bounded weights. Here, we determine a relatively precise polynomial bound $P_c(k)$ where c denotes an upper bound on the number of label changes. We view c as a constant and k as a variable that may grow to infinity. Consider a decision list over the domain $\mathcal{X}_k = \{0,1\}^k$ as given in (25). By adding redundant items if necessary, we may assume that the list contains every Boolean variable exactly once. For sake of simplicity, we apply a renumbering of indices so that the decision list looks like

$$[(z_k^{\varepsilon_k}, b_k), \ldots, (z_1^{\varepsilon_1}, b_1), b_0] \ . \tag{26}$$

Let us first assume that $\varepsilon_1 = \ldots = \varepsilon_k = 1$, i.e., all literals in the list are unnegated. We may clearly assume that $b_1 \neq b_0$ so that the first label change (from right to left) occurs between b_0 and b_1. Let $c' \leq c$ denote the total number of label changes. Let f be the Boolean function represented by (26). We will design a weight vector $v_f = (w_1, \ldots, w_k)$ and a threshold t such that (23) holds. To this end, we decompose $b_1, \ldots, b_k$ (in this order) into $c' \leq c$ maximal segments without label change. Let $I_1, \ldots, I_{c'}$ (in this order) denote the corresponding partition of $\{1, \ldots, k\}$ (so that, for instance, $1 \in I_1$ and $k \in I_{c'}$). Now we choose the weights $w_1, \ldots, w_k$ and the threshold t for the given decision list in accordance with the following policy:

- For all $j = 1, \ldots, c'$, all weights in $\{w_i : i \in I_j\}$ get the same value, say $\pm W_j$ where $W_j > 0$. More precisely, the value is W_j (resp. $-W_j$) if the label associated with the segment I_j is 1 (resp. 0).
- Let $W_1 = 1$ and, for $j = 1, \ldots, c'$, let $k_j = |I_j|$. The values $(W_j)_{j=2,\ldots,c'}$ are chosen so that the following holds:
$$W_j > k_{j-1}W_{j-1} + k_{j-3}W_{j-3} + \ldots + \begin{cases} k_1 W_1 & \text{if } j \text{ is even} \\ k_2 W_2 & \text{if } j \text{ is odd} \end{cases}$$
- The threshold t is set to $1/2$ (resp. $-1/2$) if $b_0 = 0$ (resp. $b_0 = 1$).

It is easy to verify that any weight vector and threshold chosen in accordance with this policy leads to a representation v_f of the function f associated with (26) such that (23) holds (provided that $\varepsilon_1 = \ldots = \varepsilon_k = 1$). The following is an easy-to-solve recursion for W_j leading to weights $w_1, \ldots, w_k$ that actually do respect the policy:

1. $k_0 = W_0 = 1$. For $j = 1, \ldots, c' : k_j = |I_j|$.
2. For $j = 0, \ldots, c' - 1$: $W_{j+1} = \sum_{i=0}^{j} k_i W_i$.

[5] introduced in [19]

Suppose that the weights W_j are chosen according to this recursion. We claim that, for all $j = 0, 1, \ldots, c'$,

$$W_j = \sum_{I \subseteq \{1,\ldots,j-1\}} \prod_{i \in I} k_i \ . \tag{27}$$

For $j = 0$, this is correct because the right hand-side in (27) collapses to $\prod_{i\in\emptyset} k_i = 1 = W_0$. Assume inductively that the claim is correct for W_j. Then,

$$\begin{aligned} W_{j+1} &= k_j W_j + \sum_{i=0}^{j-1} k_i W_i = (k_j+1)W_j \\ &= (k_j+1) \sum_{I \subseteq \{1,\ldots,j-1\}} \prod_{i\in I} k_i = \sum_{I \subseteq \{1,\ldots,j\}} \prod_{i\in I} k_i \ , \end{aligned}$$

which proves the claim. The method of Lagrangian multipliers yields that the largest weight $W_{c'}$, viewed as a function in $k_1, \ldots, k_{c'}$ subject to $c' \le c$ and $\sum_{j=1}^{c'} k_j = k$, is maximized for $k_{c'} = 1$ and $k_1 = \ldots = k_{c'-1} = (k-1)/(c'-1)$. Thus, we obtain the following upper bound on the weights used for the representation of the decision list:

$$\begin{aligned} \sum_{I \subseteq \{1,\ldots,c'-1\}} \prod_{i\in I} k_i &= \sum_{j=0}^{c'-1} \binom{c'-1}{j} \left(\frac{k-1}{c'-1}\right)^j \\ &= \left(1 + \frac{k-1}{c'-1}\right)^{c'-1} \end{aligned}$$

Calculus yields that the worst-case is obtained for $c' = c$. We finally remove the artificial assumption $\varepsilon_1 = \ldots = \varepsilon_k = 1$. Suppose that the parameters $\varepsilon_i \in \{0,1\}$ are not necessarily set to 1. If $\varepsilon_i = 0$, we should think of x_i in (23) as being replaced by $1 - x_i$. Multiplying out, we see that this transformation has an effect on the threshold t only, but not on the weight vector v_f. Thus, the upper bounds obtained for the absolute values of the weights remain valid. The whole discussion can be summarized as follows:

THEOREM 9. *Let $c \ge 1$ be a constant. Then, for all $r \ge 0$, the following holds. A decision list with at most k Boolean variables and with at most c label changes is linearly separable with weights whose absolute values are bounded by*

$$P_c(k) = \left(1 + \frac{k-1}{c-1}\right)^{c-1} \ .$$

7. EFFICIENCY ISSUES

Theorem 2 ignores efficiency issues. However, an inspection of its proof in Section 5 shows that efficiency is actually obtained under the following conditions:

1. Given $x \in \mathcal{X}_k$ (resp. $f \in \mathcal{F}_k$), the vector u_x (resp. v_f) of the corresponding (low-dimensional) linear arrangement $\mathcal{A}_k$ can be computed efficiently.

2. Given $f \in \mathcal{F}_k$, a random instance in $\mathcal{X}_{f,0}$ (resp. $\mathcal{X}_{f,1}$) can be efficiently generated.

The only little change in the computation of $\hat{\omega}$ (in comparison to the computation described in Section 5) is as follows. We replace the right hand-side in (16) by the corresponding empirical average taken over a sample drawn uniformly at random from $\mathcal{X}_{f,b}$. An easy application of Hoeffding's bound shows that (under the assumptions made in Theorem 2) the size of this sample can be polynomially bounded in terms of the relevant parameters. It is furthermore not hard to show that the above two conditions are satisfied by all classes discussed in Section 6.

8. CONCLUSIONS

This paper investigated to what extent randomized response schemes can combine ε-differential privacy with the conflicting goal of providing useful answers to counting queries taken from a known class $\mathcal{F}$. We introduced the notions of weak and strong usefulness and proved the following results. First, if $\mathcal{F}$ cannot be weakly SQ-learned under the uniform distribution, then ε-differential privacy rules out weak usefulness. Second, for a broad variety of classes $\mathcal{F}$ that actually can be weakly SQ-learned under the uniform distribution, we designed a randomized response scheme that provides ε-differential privacy and strong usefulness. Finally, we presented some applications. In particular, we showed that our ε-differentially private randomized response scheme is strong useful for the class of Boolean Monomials, Boolean Clauses and, more generally, for "Boolean Decision Lists with a bounded number of label-changes". The same result also applies to "Axis-Parallel Hyper-Rectangles over a discrete domain".

9. REFERENCES

[1] R. I. Arriaga and S. Vempala. An algorithmic theory of learning: Robust concepts and random projection. In *Proceedings of the 40'th Annual Symposium on the Foundations of Computer Science*, pages 616–623, 1999.

[2] S. Ben-David, N. Eiron, and H. U. Simon. Limitations of learning via embeddings in euclidean half-spaces. *Journal of Machine Learning Research*, 3:441–461, 2002.

[3] A. Blum, M. Furst, J. Jackson, M. Kearns, Y. Mansour, and S. Rudich. Weakly learning DNF and characterizing statistical query learning using Fourier analysis. In *Proceedings of the 26th Annual Symposium on Theory of Computing*, pages 253–263, 1994.

[4] A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. *Journal of the Association on Computing Machinery*, 60(2):12, 2013.

[5] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Privacy aware learning. In *Advances in Neural Information Processing Systems 25*, pages 1430–1438, 2012.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference*, pages 265–284, 2006.

[7] V. Feldman. A complete characterization of statistical query learning with applications to evolvability. In *Proceedings of the 50'th Annual Symposium on the Foundations of Computer Science*, pages 375–384, 2009.

[8] J. Forster. A linear lower bound on the unbounded error communication complexity. *Journal of Computer and System Sciences*, 65(4):612–625, 2002.

[9] J. Forster, N. Schmitt, H. U. Simon, and T. Suttorp. Estimating the optimal margins of embeddings in

euclidean half spaces. *Machine Learning*, 51(3):263–281, 2003.

[10] A. Gupta, M. Hardt, A. Roth, and J. Ullman. Privately releasing conjunctions and the statistical query barrier. *SIAM Journal on Computing*, 42(4):1494–1520, 2013.

[11] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[12] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mapping into Hilbert spaces. *Contemp. Math.*, 26:189–206, 1984.

[13] M. Kallweit and H. U. Simon. A close look to margin complexity and related parameters. In *JMLR Workshop and Conference Proceedings, Volume 19: COLT 2011*, pages 209–223, 2011.

[14] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately. *SIAM Journal on Computing*, 40(3):793–826, 2011.

[15] M. Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the Association on Computing Machinery*, 45(6):983–1006, 1998.

[16] S. Kotz, T. Kozubowski, and K. Podgorski. *The Laplace Distribution and Generalizations.* Birkhäuser, 2001.

[17] E. L. Lehmann and J. P. Romano. *Testing Statistical Hypotheses.* Springer, 2006.

[18] D. Pollard. *Convergence of Stochastic Processes.* Springer, 1984.

[19] R. Rivest. Learning decision lists. *Machine Learning*, 2(3):229–246, 1987.

[20] L. G. Valiant. Evolvability. *Journal of the Association on Computing Machinery*, 56(1):3:1–3:21, 2009.

APPENDIX

A. PROOF OF LEMMA 5

We say that z is "δ-bad" for (f,b) if $|h_{f,b}(z) - h(z)| > \delta h(z)$. Note that z is δ-bad for $(f,1)$ iff z is δ-bad for $(f,0)$ because of (10). By the Chebychef bound,

$$\begin{aligned}\Pr_{f\sim\mathcal{F}^q}[z \text{ is } \delta\text{-bad for } (f,1)] &= \Pr_{f\sim\mathcal{F}^q, b\sim B}[z \text{ is } \delta\text{-bad for } (f,b)] \\ &\le \frac{\mathrm{Var}_{f\sim\mathcal{F}^q, b\sim B}[h_{f,b}(z)]}{\delta^2 h(z)^2} \\ &\overset{(9)}{\le} \frac{U}{\delta^2}\end{aligned}$$

holds for every z. Therefore,

$$\Pr_{f\sim\mathcal{F}^q, z\sim Z(X)}[z \text{ is } \delta\text{-bad for } (f,1)] \le \frac{U}{\delta^2} \ . \tag{28}$$

Let $0 \le \beta \le 1$ be maximal such that there exists $\mathcal{F}' \subseteq \mathcal{F}$ with the following properties:

1. $\sum_{f\in\mathcal{F}'} q(f) \ge \alpha$.
2. For every $f \in \mathcal{F}'$, $\Pr_{z\sim Z(X)}[z \text{ is } \delta\text{-bad for } (f,1)] > \beta$.

The definition of β implies that

$$\Pr_{f\sim\mathcal{F}^q, z\sim Z(X)}[z \text{ is } \delta\text{-bad for } (f,1)] > \alpha\beta \ .$$

In view of (28), we may conclude that $\beta < U/(\alpha\delta^2)$. Set now $\beta_0 = U/(\alpha\delta^2)$. It follows that the probability, taken over $f \sim \mathcal{F}^q$, for

$$\Pr_{z\sim Z(X)}[z \text{ is not } \delta\text{-bad for } (f,1)] \ge 1 - \beta_0 \tag{29}$$

is at least $1-\alpha$. Consider now a fixed pair $(f,1)$ for which (29) actually holds. Let z_{bad} range over the values z that are δ-bad for $(f,1)$, and let z_{good} range over the remaining values. We are now ready for bounding the statistical difference:

$$\begin{aligned}\mathrm{SD}(Z(X_{f,1}), Z(X)) &= \frac{1}{2}\sum_z |h_{f,1}(z) - h(z)| \\ &= \frac{1}{2}\left(\sum_{z_{good}} |h_{f,1}(z) - h(z)| + \sum_{z_{bad}} |h_{f,1}(z) - h(z)|\right) \\ &\le \frac{1}{2}\left(\delta + \frac{U}{\alpha\delta^2}\right)\end{aligned}$$

Setting $\delta = (U/\alpha)^{1/3}$, we obtain the upper bound $(U/\alpha)^{1/3}$ on $\mathrm{SD}(Z(X_{f,1}), Z(X))$, as desired. □

B. PROOF OF THEOREM 6

Let $S = \sum_{i=1}^n \sum_{j=1}^d v_{i,j} Y_{i,j}$. By the Markov inequality, we get

$$\Pr[S > \delta] = \Pr\left[e^{tS} > e^{t\delta}\right] \le \frac{\mathbb{E}\left[e^{tS}\right]}{e^{t\delta}}$$

for every $t > 0$. Note that $m_S(t) = \mathbb{E}\left[e^{tS}\right]$ is the moment generating function of S and $m_Y(t) = \mathbb{E}\left[e^{tY}\right] = \frac{1}{1-(\lambda t)^2}$, defined for $t < 1/\lambda$, is the moment generating function of $Y \sim \mathrm{Lap}(\lambda)$ [16]. Suppose that $(\lambda t)^2 \le 1/2$ so that $1 - (\lambda t)^2 > \exp\left(-2(\lambda t)^2\right)$. Making use of the assumption $\|v_i\|_2 = 1$, it follows that

$$\begin{aligned}m_S(t) = \prod_{i=1}^n \prod_{j=1}^d \frac{1}{1 - v_{i,j}^2(\lambda t)^2} &< \prod_{i=1}^n \prod_{j=1}^d \frac{1}{\exp(-2v_{i,j}^2(\lambda t)^2)} \\ &= \exp\left(2(\lambda t)^2 n\right) \ .\end{aligned}$$

Setting $t = \delta/(4\lambda^2 n)$, we may conclude that

$$\Pr[S > \delta] \le \frac{m_S(t)}{e^{t\delta}} < \exp\left(2\lambda^2 t^2 n - t\delta\right) = \exp\left(-\frac{\delta^2}{8n\lambda^2}\right) \ ,$$

as desired. Note that the assumption $\delta \le \sqrt{8}\lambda n$ makes sure that the constraint $\lambda^2 t^2 \le 1/2$, a constraint that we needed in the course of our proof, actually holds for our final choice $t = \delta/(4\lambda^2 n)$. □

Setting $v_i = (1,0,\ldots,0)$ for $i = 1,\ldots,n$, Theorem 6 collapses to the following result:

COROLLARY 1. *For $i = 1,\ldots,n$, $Y_i \sim \mathrm{Lap}(\lambda)$ be i.i.d. random variables. Then, for each $\delta > 0$ such that $\delta \le \sqrt{8}\lambda n$, the following holds:*

$$\Pr\left[\sum_{i=1}^n Y_i \ge \delta\right] \le \exp\left(-\frac{\delta^2}{8n\lambda^2}\right) \tag{30}$$

Again, the assumption $\delta \le \sqrt{8}\lambda n$ is not very restrictive.

We would like to note that a result similar to Corollary 1 is found in [14]. The proof there, however, is (slightly) flawed (although the flaw can easily be fixed at the expense of a slightly weaker statement). Our proof of Theorem 6 actually builds on the proof given in [14] (except for the slightly flawed part).

C. PROOF OF THEOREM 8

We will make use of the following two results:

LEMMA 10 ([1]). *Let $u \in \mathbb{R}^r$ be arbitrary but fixed. Let $R = (R_{i,j})$ be a random $(d \times r)$-matrix such that the entries $R_{i,j}$ are i.i.d. according to the normal distribution $\mathcal{N}(0,1)$. Consider the random projection $u' := \frac{1}{\sqrt{d}}(Ru) \in \mathbb{R}^d$. Then the following holds for every constant $\gamma > 0$:*

$$\Pr\left[\left|\|u'\|_2^2 - \|u\|_2^2\right| \geq \gamma\|u\|_2^2\right] \leq 2e^{-\gamma^2 d/8}$$

LEMMA 11 ([2]). *Let $v, x \in \mathbb{R}^r$ be arbitrary but fixed. Let R be the same random $(d \times r)$-matrix and let $u \mapsto u'$ be the same random projection as in Lemma 10. Then the following holds for every constant $\gamma > 0$:*

$$\Pr\left[\left|\langle v', x'\rangle - \langle v, x\rangle\right| \geq \frac{\gamma}{2}(\|v\|_2^2 + \|x\|_2^2)\right] \leq 4e^{-\gamma^2 d/8}$$

We are now prepared for the proof of Theorem 8. Let $M := |\mathcal{X}|$ be the number of instances in the universe, and let $N := |\mathcal{F}|$ be the number of counting queries in $\mathcal{F}$. Suppose that $\mathcal{A}$ is an r-dimensional arrangement for $\mathcal{F}$ (resp. for the corresponding sign matrix A) such that $\gamma := \tilde{\gamma}_{min}(A|\mathcal{A}) > 0$. Let $\mathcal{A}'$ be the d-dimensional arrangement for $\mathcal{F}$ that results from $\mathcal{A}$ by randomly projecting every vector $u \in \{u_x : x \in \mathcal{X}\}$ (resp. every vector $v \in \{v_f : f \in \mathcal{F}\}$) to u' (resp. to v'). Recall that the vectors u_x, v_f have unit Euclidean length. Let p_1 denote the probability for

$$\exists w \in \{u_x : x \in \mathcal{X}\} \cup \{v_f : f \in \mathcal{F}\} : \|w'\|_2 \geq 2 \ . \qquad (31)$$

According to Lemma 10, applied with 1 in the role of γ and combined with the union bound, we get $p_1 \leq 2(M+N)e^{-d/8}$. For every $f \in \mathcal{F}$, consider the vector

$$x_f = \frac{1}{2} \cdot \sum_{b=0,1} \left(\frac{1}{|\mathcal{X}_{f,b}|} \sum_{x \in \mathcal{X}_{f,b}} A_{x,f} u_x \right) \ .$$

Note that $\|x_f\|_2 \leq 1$. Let p_2 denote the probability for

$$\exists f \in \mathcal{F} : \left|\langle v'_f, x'_f\rangle - \langle v_f, x_f\rangle\right| \geq \frac{\gamma}{2} \ . \qquad (32)$$

Note that $\langle v_f, x_f\rangle = \tilde{\gamma}_f(A|\mathcal{A}) \geq \gamma$ for all $f \in \mathcal{F}$. Moreover,

$$\frac{\gamma}{2} = 2 \cdot \frac{\gamma}{4} \geq \left(\|v_f\|_2^2 + \|x_f\|_2^2\right) \cdot \frac{\gamma}{4} \ .$$

According to Lemma 11, applied with $\gamma/4$ in the role of γ and combined with the union bound, we get

$$p_2 \leq 4Ne^{-\gamma^2 d/128}.$$

As an easy calculation shows, condition (22) implies that, with a probability of at least $1 - \beta$, none of the events (31), (32) occurs. If none of these events occurs, then

$$\tilde{\gamma}_{min}(A|\mathcal{A}') = \min_{f \in \mathcal{F}} \tilde{\gamma}_f(A|\mathcal{A}') \geq (1/4)(\gamma/2) = \gamma/8 \ ,$$

where the scaling factor 1/4 takes into account that the vectors in the arrangement $\mathcal{A}'$ (with an Euclidean length of at most 2) must still be normalized so as to have unit Euclidean length. □