CONCENTRATION BOUNDS FOR HIGH SENSITIVITY FUNCTIONS THROUGH DIFFERENTIAL PRIVACY

KOBBI NISSIM AND URI STEMMER

Georgetown University
e-mail address: kobbi.nissim@georgetown.edu

Ben-Gurion University
e-mail address: u@uri.co.il

ABSTRACT. A new line of work, starting with Dwork et al. (STOC, 2015), demonstrates how differential privacy can be used as a mathematical tool for guaranteeing generalization in adaptive data analysis. Specifically, if a differentially private analysis is applied on a sample S of i.i.d. examples to select a low-sensitivity function f, then w.h.p. $f(S)$ is close to its expectation, even though f is being chosen adaptively, i.e., based on the data.

Very recently, Steinke and Ullman observed that these generalization guarantees can be used for proving concentration bounds in the non-adaptive setting, where the low-sensitivity function is fixed beforehand. In particular, they obtain alternative proofs for classical concentration bounds for low-sensitivity functions, such as the Chernoff bound and McDiarmid's Inequality. In this work, we extend this connection between differential privacy and concentration bounds, and show that differential privacy can be used to prove concentration of functions that are not low-sensitivity.

1. INTRODUCTION

A new line of work, starting with Dwork et al. [2015b], demonstrates how differential privacy [Dwork et al., 2006b] can be used as a mathematical tool for guaranteeing statistical validity in data analysis. Specifically, if a differentially private analysis is applied on a sample S of i.i.d. examples to select a low-sensitivity function f, then w.h.p. $f(S)$ is close to its expectation, even when f is being chosen based on the data. Dwork et al. [2015b] showed how to utilize this connection for the task of answering *adaptively chosen* queries w.r.t. an unknown distribution using i.i.d. samples from it.

To make the setting concrete, consider a data analyst interested in learning properties of an unknown distribution $\mathcal{D}$. The analyst interacts with the distribution $\mathcal{D}$ via a *data curator* $\mathcal{A}$ holding a database S containing n i.i.d. samples from $\mathcal{D}$. The interaction is adaptive, where at every round the analyst specifies a query $q : X^n \to \mathbb{R}$ and receives an answer $a_q(S)$ that (hopefully) approximates $q(\mathcal{D}^n) \triangleq \mathbb{E}_{S' \sim \mathcal{D}^n}[q(S')]$. As the analyst chooses

its queries based on previous interactions with the data, we run the risk of overfitting if $\mathcal{A}$ simply answers every query with its empirical value on the sample S. However, if $\mathcal{A}$ is a differentially private algorithm then the interaction would not lead to overfitting:

Theorem 1.1 ([Dwork et al., 2015b, Bassily et al., 2016], informal)**.** *A function $f : X^n \to \mathbb{R}$ has sensitivity λ if $|f(S) - f(S')| \leq \lambda$ for every pair $S, S' \in X^n$ differing in only one entry. Define $f(\mathcal{D}^n) \triangleq \underset{S' \sim \mathcal{D}^n}{\mathbb{E}} [f(S')]$. Let $\mathcal{A} : X^n \to \mathcal{F}_\lambda$ be (ε, δ)-differentially private where $\mathcal{F}_\lambda$ is the class of λ-sensitive functions, and $n \geq \frac{1}{\varepsilon^2} \log(\frac{4\varepsilon}{\delta})$. Then for every distribution $\mathcal{D}$ on X,*

$$\Pr_{\substack{S \sim \mathcal{D}^n \\ f \leftarrow \mathcal{A}(S)}} [|f(S) - f(\mathcal{D}^n)| \geq 18\varepsilon\lambda n] < \frac{\delta}{\varepsilon}.$$

In words, if $\mathcal{A}$ is a differentially private algorithm operating on a database containing n i.i.d. samples from the distribution $\mathcal{D}$, then $\mathcal{A}$ cannot (with significant probability) identify a low-sensitivity function that behaves differently on the sample S and on $\mathcal{D}^n$.

Steinke and Ullman [2017] observed that Theorem 1.1 gives alternative proofs for classical concentration bounds for low-sensitivity functions, such as the Chernoff bound and McDiarmid's Inequality: Fix a function $f : X^n \to \mathbb{R}$ with sensitivity λ and consider the trivial mechanism $\mathcal{A}_f$ that ignores its input and always outputs f. Such a mechanism is (ε, δ)-differentially private for any choice of $\varepsilon, \delta \geq 0$ and hence Theorem 1.1 yields (up to constants) McDiarmid's Inequality:

$$\Pr_{S \sim \mathcal{D}^n} [|f(S) - f(\mathcal{D}^n)| \geq 18\varepsilon\lambda n] < \frac{\delta}{\varepsilon} = 2^{-\Omega(\varepsilon^2 \cdot n)}, \tag{1.1}$$

where the last equality follows by setting $n = \frac{1}{\varepsilon^2} \log(\frac{4\varepsilon}{\delta})$.

In light of this result it is natural to ask if similar techniques yield concentration bounds for more general families of functions, in particular functions that are not low-sensitivity. In this work we derive conditions under which this is the case.

1.1. **Differential Privacy and Max-Information.** Let $\mathcal{D}$ be a fixed distribution over a domain X, and consider a family of functions mapping databases in X^n to the reals, such that for every function f in the family we have that $|f(S) - f(\mathcal{D}^n)|$ is small w.h.p. over $S \sim \mathcal{D}^n$. Specifically,

$$\mathcal{F}_{\alpha,\beta}(\mathcal{D}) = \left\{ \ f : X^n \to \mathbb{R} \quad : \quad \Pr_{S \sim \mathcal{D}^n}[|f(S) - f(\mathcal{D}^n)| > \alpha] \leq \beta \ \right\}.$$

That is, for every function $f \in \mathcal{F}_{\alpha,\beta}(\mathcal{D})$ we have that its empirical value over a sample $S \sim \mathcal{D}^n$ is α-close to its expected value w.p. $1 - \beta$. Now consider a differentially private algorithm $\mathcal{A} : X^n \to \mathcal{F}_{\alpha,\beta}(\mathcal{D})$ that takes a database and returns a function from $\mathcal{F}_{\alpha,\beta}(\mathcal{D})$. What can we say about the difference $|f(S) - f(\mathcal{D}^n)|$ when f is chosen by $\mathcal{A}(S)$ based on the sample S itself? (Observe that f is not necessarily a low-sensitivity function.)

Using the notion of *max-information*, Dwork et al. [2015a] showed that if β is small enough, then w.h.p. the difference remains small. Informally, they showed that if $\mathcal{A}$ is differentially private, then

$$\Pr_{\substack{S \sim \mathcal{D}^n \\ f \leftarrow \mathcal{A}(S)}} [|f(S) - f(\mathcal{D}^n)| > \alpha] \lesssim \beta \cdot e^{\varepsilon^2 \cdot n}.$$

So, if $\mathcal{A}$ is a differentially private algorithm that ranges over functions which are very concentrated around their expected value (i.e., $\beta < e^{-\varepsilon^2 n}$), then $|f(S) - f(\mathcal{D}^n)|$ remains small (w.h.p.) even when f is chosen by $\mathcal{A}(S)$ based on the sample S. The results of this article provide another piece of this puzzle, as we show that (a variant of) differential privacy can in some cases be used to prove that a function f is in $\mathcal{F}_{\alpha,\beta}(\mathcal{D})$, even for functions for which differential privacy does not guarantee generalization under adaptive selection.

1.2. **Our Results.**

Notation. Throughout this article we use the convention that $f(\mathcal{D}^n)$ is the expected value of the function f over a sample containing n i.i.d. elements drawn according to the distribution $\mathcal{D}$. That is, $f(\mathcal{D}^n) \triangleq \mathbb{E}_{S \sim \mathcal{D}^n} [f(S)]$.

Fix a function $f : X^n \to \mathbb{R}$, let $\mathcal{D}$ be a distribution over X, and let $S \sim \mathcal{D}^n$. Our goal is to bound the probability that $|f(S) - f(\mathcal{D}^n)|$ is large by some (hopefully) easy-to-analyze quantity. To intuit our result, consider for example what we get by a simple application of Markov's Inequality:

$$\Pr_{S \sim \mathcal{D}^n}[|f(S) - f(\mathcal{D}^n)| > \lambda] \leq \frac{1}{\lambda} \cdot \mathbb{E}_{S \sim \mathcal{D}^n} \left[\mathbb{1}_{|f(S) - f(\mathcal{D}^n)| > \lambda} \cdot |f(S) - f(\mathcal{D}^n)| \right]. \qquad (1.2)$$

We show that, using differential privacy, the term $|f(S) - f(\mathcal{D}^n)|$ in the expectation can replaced with $|f(S \cup \{x\}) - f(S \cup \{y\})|$, which is sometimes easier to analyze. Our main contribution – Theorem 1.2 – has two parts. We note that a variant of Part 1 can be obtained by applying Azuma's inequality, and that our emphasis here is on the new proof technique. We are not aware of an alternative proof for Part 2, which is needed for some of our applications (see Section B for more details).

Theorem 1.2 (Part 1). *Let $n, \Delta, \lambda, \varepsilon > 0$ be s.t. $n \geq O\left(\frac{1}{\varepsilon \cdot \min\{1,\varepsilon\}} \log(\frac{\lambda \cdot \min\{1,\varepsilon\}}{\Delta}) \right)$. Let $\mathcal{D}$ be a distribution over a domain X, let $f : X^n \to \mathbb{R}$ be s.t.*

$$\forall i \in [n], \quad \mathbb{E}_{\substack{S \sim \mathcal{D}^n \\ z \sim \mathcal{D}}} \left[\mathbb{1}_{|f(S) - f(S^{(i \leftarrow z)})| > \lambda} \cdot \left| f(S) - f\left(S^{(i \leftarrow z)} \right) \right| \right] \leq \Delta, \qquad (1.3)$$

where $S^{(i \leftarrow z)}$ is the same as S except that the ith element is replaced with z. Then,

$$\Pr_{S \sim \mathcal{D}^n} [|f(S) - f(\mathcal{D}^n)| \geq 18\varepsilon\lambda n] < \frac{14\Delta}{\varepsilon\lambda}.$$

Observe that for a λ-sensitive function f, we have that the expectation in Equation (1.3) is zero, so the statement holds for every choice of $\Delta > 0$ and $n \geq O\left(\frac{1}{\varepsilon^2} \log(\frac{\lambda\varepsilon}{\Delta}) \right)$, resulting in McDiarmid's Inequality (Equation (1.1)). Intuitively, Theorem 1.2 states that in order to obtain a high probability bound on $|f(S) - f(\mathcal{D}^n)|$ it suffices to analyze the "expectation of the tail" of $\left| f(S) - f\left(S^{(i \leftarrow z)} \right) \right|$, as a function of the starting point λ.

We also show that the above bound can be improved whenever the "expectation of the head" of $\left| f(S) - f\left(S^{(i \leftarrow z)} \right) \right|$ is smaller than λ. Specifically,

Theorem 1.2 (Part 2). *Let $n, \Delta, \lambda, \tau, \varepsilon > 0$ be s.t. $n \geq O\left(\frac{\lambda}{\varepsilon \cdot \min\{1,\varepsilon\}\tau} \log(\frac{\tau \cdot \min\{1,\varepsilon\}}{\Delta}) \right)$, and denote $T = \left\lfloor \frac{(e^\varepsilon - e^{-\varepsilon})\tau}{7\Delta} \right\rfloor$. Let $\mathcal{D}$ be a distribution over a domain X, and let $f : X^n \to \mathbb{R}$. If,*

in addition to (1.3), for every $i \in [n]$ we have

$$\mathop{\mathbb{E}}_{S_1,\ldots,S_T \sim \mathcal{D}^n} \max_{t \in [T]} \mathop{\mathbb{E}}_{y,z \sim \mathcal{D}} \left[\mathbb{1}_{\left| f(S_t^{(i \leftarrow y)}) - f\left(S_t^{(i \leftarrow z)}\right) \right| \leq \lambda} \cdot \left| f(S_t^{(i \leftarrow y)}) - f\left(S_t^{(i \leftarrow z)}\right) \right| \right] \leq \tau, \quad (1.4)$$

then

$$\Pr_{S \sim \mathcal{D}^n} \left[|f(S) - f(\mathcal{D}^n)| \geq 18\varepsilon\tau n \right] < \frac{14\Delta}{\varepsilon\tau}.$$

To see the potential of Part 2, consider again the case where f is a λ-sensitive function, and recall that in that case Part 1 results in McDiarmid's Inequality (Equation (1.1)). Now, if we can show that condition (1.4) holds for some $\tau < \lambda$, then we would get the following stronger bound:

$$\Pr_{S \sim \mathcal{D}^n} \left[|f(S) - f(\mathcal{D}^n)| \geq 18\varepsilon\tau n \right] < \frac{14\Delta}{\varepsilon\tau} = 2^{-\Omega(\frac{\tau}{\lambda}\varepsilon^2 \cdot n)}.$$

The advantage over McDiarmid's Inequality is that the bound on $|f(S) - f(\mathcal{D}^n)|$ is potentially smaller ($\varepsilon\tau n$ as opposed to $\varepsilon\lambda n$).

Remark 1.3. *We remark that Part 2 of Theorem 1.2 would become stronger if the parameter T could be made smaller (ideally we would want to have $T = 1$). We do not know if it is possible to reduce T. We also remark that requirement (1.4) can be replaced with*

$$\forall S \in X^n, \ \ \forall i \in [n], \quad \mathop{\mathbb{E}}_{y,z \sim \mathcal{D}} \left[\mathbb{1}_{\left| f(S^{(i \leftarrow y)}) - f\left(S^{(i \leftarrow z)}\right) \right| \leq \lambda} \cdot \left| f(S^{(i \leftarrow y)}) - f\left(S^{(i \leftarrow z)}\right) \right| \right] \leq \tau. \quad (1.5)$$

Indeed, whenever (1.5) holds we have that the inner expectation in (1.4) is at most τ, and hence condition (1.4) holds.

In Section 4 we demonstrate how Theorem 1.2 can be used in proving a variety of concentration bounds, such as a high probability bound on $|f(S) - f(\mathcal{D}^n)|$ for Lipschitz functions. In addition we show that Theorem 1.2 can be applied to bound the probability that the number of triangles in a random graph significantly exceeds the expectation. Consider a random graph on N vertices defined by drawing each edge independently with probability $p = o(1)$. The expected number of triangles is $\Theta(N^3 p^3) = o(N)$. Note that, in worst case, adding a single edge to the graph can increase the number of triangles by $N - 1$, a quantity larger than the expected number of triangles. With such a high (worst-case) Lipschitz constant, a direct application of McDiarmid's Inequality would not result in meaningful bound. However, using Theorem 1.2 only requires arguing about the *expected* increase in the number of triangles when resampling a single edge, which is much smaller.

In Section 5 we construct an example for a differentially private algorithm that, given a sample $S \sim \mathcal{D}^n$, identifies (w.h.p.) a function f such that $|f(S)|$ is large, even though f is highly concentrated around zero (on a fresh sample from the underlying distribution). We interpret this result as an indication that differential privacy can be used to prove concentration bounds even for functions for which it does not guarantee generalization under adaptive selection.

2. Preliminaries

2.1. **Differential Privacy.** Our results rely on a number of basic facts about differential privacy. An algorithm operating on databases is said to preserve differential privacy if a change of a single record of the database does not significantly change the output distribution of the algorithm. Formally:

Definition 2.1. Databases $S \in X^n$ and $S' \in X^n$ over a domain X are called *neighboring* if they differ in exactly one entry.

Definition 2.2 (Differential Privacy [Dwork et al., 2006b,a]). A randomized algorithm $\mathcal{A} : X^n \to Y$ is (ε, δ)-*differentially private* if for all neighboring databases $S, S' \in X^n$, and for every set of outputs $T \subseteq Y$, we have

$$\Pr[\mathcal{A}(S) \in T] \leq e^\varepsilon \cdot \Pr[\mathcal{A}(S') \in T] + \delta.$$

The probability is taken over the random coins of $\mathcal{A}$.

2.2. **The Exponential Mechanism.** We next describe the exponential mechanism of McSherry and Talwar [2007].

Definition 2.3 (Sensitivity). The *sensitivity* (or *global sensitivity*) of a function $f : X^n \to \mathbb{R}$ is the smallest λ such that for every neighboring $S, S' \in X^n$, we have $|f(S) - f(S')| \leq \lambda$. We use the term "$\lambda$-sensitive function" to mean a function of sensitivity $\leq \lambda$.

Let X be a domain and H a set of solutions. Given a database $S \in X^*$, the exponential mechanism privately chooses a "good" solution h out of the possible set of solutions H. This "goodness" is quantified using a *quality function* that matches solutions to scores.

Definition 2.4 (Quality function). A *quality function* is a function $q : X^* \times H \to \mathbb{R}$ that maps a database $S \in X^*$ and a solution $h \in H$ to a real number, identified as the score of the solution h w.r.t. the database S.

Given a quality function q and a database S, the goal is to choose a solution h approximately maximizing $q(S, h)$. The exponential mechanism chooses a solution probabilistically, where the probability mass that is assigned to each solution h increases exponentially with its quality $q(S, h)$:

The Exponential Mechanism
Input: privacy parameter $\varepsilon > 0$, finite solution set H, database $S \in X^n$, and a λ-sensitive quality function q.

(1) Randomly choose $h \in H$ with probability $\frac{\exp\left(\frac{\varepsilon}{2\lambda} \cdot q(S,h)\right)}{\sum_{h' \in H} \exp\left(\frac{\varepsilon}{2\lambda} \cdot q(S,h')\right)}$.

(2) Output h.

Theorem 2.5 ([McSherry and Talwar, 2007]). *(i) The exponential mechanism is $(\varepsilon, 0)$-differentially private. (ii) The expected quality of the output is large, specifically,*

$$\mathop{\mathbb{E}}_{h \in_R H}[q(S,h)] \geq \max_{h \in H}\{q(S,h)\} - \frac{2\lambda}{\varepsilon} \cdot \ln |H|.$$

We remark that the utility guarantees of the exponential mechanism are stronger than the above statement, in the sense that the quality of the output is large *with high probability*. However, the formulation of Theorem 2.5 will be more convenient for our application of the mechanism (see, e.g., [Bassily et al., 2016] for an analysis).

2.3. **Chernoff and Hoeffding Bounds.** Let $X_1, \ldots, X_n$ be independent random variables where $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$ for some $0 < p < 1$. Clearly, $\mathbb{E}[\sum_{i=1}^{n} X_i] = pn$. Chernoff and Hoeffding bounds show that the sum is concentrated around this expected value:

$$\Pr\left[\sum_{i=1}^{n} X_i > (1 + \delta)pn\right] \leq \exp\left(-pn\delta^2/3\right) \quad \text{for } 0 < \delta \leq 1,$$

$$\Pr\left[\sum_{i=1}^{n} X_i < (1 - \delta)pn\right] \leq \exp\left(-pn\delta^2/2\right) \quad \text{for } 0 < \delta < 1,$$

$$\Pr\left[\left|\sum_{i=1}^{n} X_i - pn\right| > \delta\right] \leq 2\exp\left(-2\delta^2/n\right) \quad \text{for } \delta \geq 0.$$

The first two inequalities are known as the multiplicative Chernoff bounds [Chernoff, 1952], and the last inequality is known as the Hoeffding bound [Hoeffding, 1963]. The next theorem states that the Chernoff bound above is tight up to constant factors in the exponent.

Theorem 2.6 (Tightness of Chernoff bound [Klein and Young, 2015]). *Let $0 < p, \delta \leq \frac{1}{2}$, and let $n \geq \frac{3}{\delta^2 p}$. Let $X_1, \ldots, X_n$ be independent random variables where $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$. Then,*

$$\Pr\left[\sum_{i=1}^{n} X_i \leq (1 - \delta)pn\right] \geq \exp(-9\delta^2 pn),$$

$$\Pr\left[\sum_{i=1}^{n} X_i \geq (1 + \delta)pn\right] \geq \exp(-9\delta^2 pn).$$

3. Concentration Bounds via Differential Privacy

In this section we show how the concept of differential privacy can be used to derive conditions under which a function f and a distribution $\mathcal{D}$ satisfy that $|f(S) - f(\mathcal{D}^n)|$ is small w.h.p. when $S \sim \mathcal{D}^n$. Our proof technique builds on the proof of Bassily et al. [2016] for the generalization properties of a differentially private algorithm that outputs a low-sensitivity function. The proof consists of two steps:

(1) Let $S_1, \ldots, S_T$ be T independent samples from $\mathcal{D}^n$ (each containing n i.i.d. samples from $\mathcal{D}$). Let $\mathcal{A}$ be selection procedure that, given $S_1, \ldots, S_T$, chooses an index $t \in [T]$ with the goal of maximizing $|f(S_t) - f(\mathcal{D}^n)|$. We show that if $\mathcal{A}$ satisfies (a variant of) differential privacy then, under some conditions on the function f and the distribution $\mathcal{D}$, the expectation of $|f(S_t) - f(\mathcal{D}^n)|$ is bounded. That is, if $\mathcal{A}$ is differentially private, then its ability to identify a "bad" index t with large $|f(S_t) - f(\mathcal{D}^n)|$ is limited.

(2) We show that if there is a noticeable probability (over $S \sim \mathcal{D}^n$) that $|f(S) - f(\mathcal{D}^n)|$ is large, then it is possible to construct an algorithm $\mathcal{A}$ satisfying (a variant of) differential privacy that contradicts our expectation bound.

We begin with a few definitions.

3.1. Definitions.

Notations. We use $\vec{S} \in (X^n)^T$ to denote a *multi*-database consisting of T databases of size n over X. Given a distribution $\mathcal{D}$ over a domain X we write $\vec{S} \sim \mathcal{D}^{nT}$ to denote a multi-database sampled i.i.d. from $\mathcal{D}$.

Definition 3.1. Fix a function $f : X^n \to \mathbb{R}$ mapping databases of size n over a domain X to the reals. We say that two multi-databases $\vec{S} = (S_1, \ldots, S_T) \in (X^n)^T$ and $\vec{S}' = (S_1', \ldots, S_T') \in (X^n)^T$ are (f, λ)-*neighboring* if for all $1 \le i \le T$ we have that

$$|f(S_i) - f(S_i')| \le \lambda.$$

Definition 3.2 $((\varepsilon, (f, \lambda))$-differential privacy$)$. Let $M : (X^n)^T \to Y$ be a randomized algorithm that operates on T databases of size n from X. For a function $f : X^n \to \mathbb{R}$ and parameters $\varepsilon, \lambda \ge 0$, we say that M is $(\varepsilon, (f, \lambda))$-*differentially private* if for every set of outputs $F \in Y$ and for every (f, λ)-neighboring $\vec{S}, \vec{S}' \in (X^n)^T$ it holds that

$$\Pr[M(\vec{S}) \in F] \le e^{\varepsilon} \cdot \Pr[M(\vec{S}') \in F].$$

Claim 3.3. *Fix a function $f : X^n \to \mathbb{R}$ and parameters $\varepsilon \le 1$ and $\lambda \ge 0$. If $M : (X^n)^T \to Y$ is $(\varepsilon, (f, \lambda))$-differentially private then for every (f, λ)-neighboring databases $\vec{S}, \vec{S}' \in (X^n)^T$ and every function $h : Y \to \mathbb{R}$ we have that*

$$\mathop{\mathbb{E}}_{y \leftarrow M(\vec{S})} [h(y)] \le \mathop{\mathbb{E}}_{y \leftarrow M(\vec{S}')} [h(y)] + 4\varepsilon \cdot \mathop{\mathbb{E}}_{y \leftarrow M(\vec{S}')} [|h(y)|].$$

Claim 3.3 follows from basic arguments in differential privacy. The proof appears in the appendix for completeness.

3.2. Multi Sample Expectation Bound.

The proof of Theorem 1.2 contains somewhat unwieldy notations. For readability, we present here a restricted version of the theorem, tailored to the case where the function f computes the sample sum, which highlights most of the ideas in the proof. The full proof of Theorem 1.2 is included in the appendix.

Notation. Given a sample $S \in X^n$, we use $\bar{f}(S)$ to denote the sample sum, i.e., $\bar{f}(S) = \sum_{x \in S} x$.

Lemma 3.1 (Simplified Expectation Bound). *Let $\mathcal{D}$ be a distribution over a domain X such that $\mathop{\mathbb{E}}_{x \sim \mathcal{D}}[x] = 0$ and $\mathop{\mathbb{E}}_{x \sim \mathcal{D}}[\mathbb{1}_{\{|x| > 1\}} \cdot |x|] \le \Delta$. Fix $0 < \varepsilon \le 1$, and let $\mathcal{A} : (X^n)^T \to [T]$ be an $(\varepsilon, (\bar{f}, 1))$-differentially private algorithm that operates on T databases of size n from X, and outputs an index $1 \le t \le T$. Then*

$$\left| \mathop{\mathbb{E}}_{\substack{\vec{S} \sim \mathcal{D}^{nT} \\ t \leftarrow \mathcal{A}(\vec{S})}} [\bar{f}(S_t)] \right| \le 4\varepsilon n + 2nT\Delta.$$

Proof. We denote $\vec{S} = (S_1, \ldots, S_T)$, where every S_t is itself a vector $S_t = (x_{t,1}, \ldots, x_{t,n})$. We have:

$$\mathop{\mathbb{E}}_{\substack{\vec{S} \sim \mathcal{D}^{nT} \\ t \leftarrow \mathcal{A}(\vec{S})}} \left[\bar{f}(S_t) \right] = \sum_{i \in [n]} \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S})} [x_{t,i}]$$

$$= \sum_{i \in [n]} \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| \leq 1 \right\} \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S})} [x_{t,i}] + \mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| > 1 \right\} \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S})} [x_{t,i}] \right].$$
$$(3.1)$$

In the case where $\max_{m \in [T]} |x_{m,i}| > 1$ we replace the expectation over $t \leftarrow \mathcal{A}(\vec{S})$ with the deterministic choice for the maximal t (this makes the expression larger). When $\max_{m \in [T]} |x_{m,i}| \leq 1$ we can use the privacy guarantees of algorithm $\mathcal{A}$. Given a multi-sample $\vec{S} \in (X^n)^T$ we use $\vec{S}_{-i}$ to denote a multi-sample identical to $\vec{S}$, except that the ith element of *every* sub-sample is replaced with 0. Using Claim 3.3 we get

$$(3.1) \leq \sum_{i \in [n]} \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| \leq 1 \right\} \cdot \left(\mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}_{-i})} [x_{t,i}] + 4\varepsilon \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}_{-i})} [|x_{t,i}|] \right) \right.$$

$$\left. + \mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| > 1 \right\} \cdot \max_{m \in [T]} |x_{m,i}| \right]$$

$$\leq 4\varepsilon n + \sum_{i \in [n]} \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| \leq 1 \right\} \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}_{-i})} [x_{t,i}] \right.$$

$$\left. + \mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| > 1 \right\} \cdot \max_{m \in [T]} |x_{m,i}| \right] \qquad (3.2)$$

We next want to remove the first indicator function. This is useful as without it, the expectation of a fresh example from $\mathcal{D}$ is zero. To that end we add and subtract the expression $\mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| > 1 \right\} \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}_{-i})} [x_{t,i}]$ to get (after replacing again $\mathbb{E}_t$ with $\max_t$)

$$(3.2) \leq 4\varepsilon n + \sum_{i \in [n]} \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}_{-i})} [x_{t,i}] + 2 \cdot \mathbb{1} \left\{ \max_{m \in [T]} |x_{m,i}| > 1 \right\} \cdot \max_{m \in [T]} |x_{m,i}| \right]$$

$$\leq 4\varepsilon n + 2 \sum_{i \in [n]} \sum_{m \in [T]} \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\mathbb{1} \left\{ |x_{m,i}| > 1 \right\} \cdot |x_{m,i}| \right]$$

$$\leq 4\varepsilon n + 2nT\Delta.$$

$\square$

3.3. **Multi Sample Amplification.**

Theorem 3.4 (Simplified High Probability Bound). *Let $\mathcal{D}$ be a distribution over a domain X such that $\mathop{\mathbb{E}}_{x \sim \mathcal{D}} [x] = 0$. Let $\Delta \geq 0$ be such that $\mathop{\mathbb{E}}_{x \sim \mathcal{D}} \left[\mathbb{1}_{\{|x|>1\}} \cdot |x| \right] \leq \Delta$. Fix $1 \geq \varepsilon \geq$*

$\sqrt{\frac{1}{n} \ln\left(\frac{2\varepsilon}{\Delta}\right)}$. *We have that*

$$\Pr_{S \sim \mathcal{D}^n} \left[|\bar{f}(S)| \geq 30\varepsilon n\right] < \frac{\Delta}{\varepsilon}.$$

For example, observe that if $\mathcal{D}$ is uniform on $\{\pm 1\}$ then the condition in the theorem holds for every choice of $\Delta > 0$ (since the expectation is zero). Hence, by setting $\varepsilon = \sqrt{\frac{1}{n} \ln\left(\frac{2\varepsilon}{\Delta}\right)}$ we get $\Pr_{S \sim \mathcal{D}^n} \left[|\bar{f}(S)| \geq 30\varepsilon n\right] < 2e^{-n\varepsilon^2}$ for every $0 \leq \varepsilon \leq 1$. We next present the proof idea of Theorem 3.4. Any informalities made herein are removed in Section A.

Proof sketch. We only analyze the probability that $\bar{f}(S)$ is large. The analysis is symmetric for when $\bar{f}(S)$ is small. Assume towards contradiction that with probability at least $\frac{\Delta}{2\varepsilon}$ over $S \sim \mathcal{D}^n$ we have that $\bar{f}(S) \geq 30\varepsilon n$. We now construct the following algorithm $\mathcal{B}$ that contradicts our expectation bound.

Algorithm 1: $\mathcal{B}$

Input: T databases of size n each: $\vec{S} = (S_1, \ldots, S_T)$, where $T \triangleq \lfloor 2\varepsilon/\Delta \rfloor$.

(1) For $i \in [T]$, define $q(\vec{S}, i) = \bar{f}(S_i)$.

(2) Sample $t^* \in [T]$ with probability proportional to $\exp\left(\frac{\varepsilon}{2} q(\vec{S}, t)\right)$.

Output: t^*.

The fact that algorithm $\mathcal{B}$ is $(\varepsilon, (\bar{f}, 1))$-differentially private follows from the standard analysis of the Exponential Mechanism of McSherry and Talwar [2007]. The analysis appears in the full version of this proof (Section A) for completeness.

Now consider applying $\mathcal{B}$ on a multi-database $\vec{S} = (S_1, \ldots, S_T)$ containing i.i.d. samples from $\mathcal{D}$. By our assumption on $\mathcal{D}$, for every t we have that $\bar{f}(S_t) \geq 30\varepsilon n$ with probability at least $\frac{\Delta}{2\varepsilon}$. By our choice of $T = \lfloor 2\varepsilon/\Delta \rfloor$, we therefore get

$$\Pr_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max_{t \in [T]} \left\{\bar{f}(S_t)\right\} \geq 30\varepsilon n\right] \geq 1 - \left(1 - \frac{\Delta}{2\varepsilon}\right)^T \geq \frac{1}{2}.$$

The probability is taken over the random choice of the examples in $\vec{S}$ according to $\mathcal{D}$. Had it been the case that the random variable $\max_{t \in [T]} \left\{\bar{f}(S_t)\right\}$ is non-negative, we could have used Markov's inequality to get

$$\mathbb{E}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max_{t \in [T]} \left\{q(\vec{S}, t)\right\}\right] = \mathbb{E}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max_{t \in [T]} \left\{\bar{f}(S_t)\right\}\right] \geq 15\varepsilon n. \tag{3.3}$$

Even though it is not the case that $\max_{t \in [T]} \left\{\bar{f}(S_t)\right\}$ is non-negative, we now proceed as if Equation (3.3) holds. As described in the full version of this proof (Section A), this technical issue has an easy fix. So, in expectation, $\max_{t \in [T]} \left(q(\vec{S}, t)\right)$ is large. In order to contradict the expectation bound of Lemma 3.1, we need to show that this is also the case for the index t^* that is sampled in Step 2. Indeed, by the properties of the exponential mechanism (Theorem 2.5) we have that

$$\mathbb{E}_{t^* \in_R [T]}[q(\vec{S}, t^*)] = \mathbb{E}_{t^* \in_R [T]} \left[\bar{f}(S_{t^*})\right] \geq \max_{t \in [T]} \left\{\bar{f}(S_t)\right\} - \frac{2}{\varepsilon} \ln(T).$$

Taking the expectation also over $\vec{S} \sim \mathcal{D}^{nT}$ we get that

$$
\mathop{\mathbb{E}}_{\substack{\vec{S} \sim \mathcal{D}^{nT} \\ t^* \leftarrow \mathcal{B}(\vec{S})}} \left[\bar{f}(S_{t^*}) \right] \geq \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max_{t \in [T]} \left\{ \bar{f}(S_t) \right\} \right] - \frac{2}{\varepsilon} \ln(T)
$$

$$
\geq 15 \varepsilon n - \frac{2}{\varepsilon} \ln(T).
$$

This contradicts Lemma 3.1 whenever $\varepsilon > \sqrt{\frac{1}{n} \ln(T)} = \sqrt{\frac{1}{n} \ln(2\varepsilon/\Delta)}$. □

4. Applications

In this section we demonstrate how Theorem 1.2 can be used in proving a variety of concentration bounds.

4.1. **Example: Subgaussian Diameter and Beyond.** Recall that for a low-sensitivity function f, one could use McDiarmid's Inequality to obtain a high probability bound on the difference $|f(S) - f(\mathcal{D}^n)|$, and this bound is *distribution-independent*. That is, the bound does not depend on $\mathcal{D}$. Over the last few years, there has been some work on providing distribution-dependent refinements to McDiarmid's Inequality, that hold even for functions with high worst-case sensitivity, but with low "average-case" sensitivity, where "average" is with respect to the underlying distribution $\mathcal{D}$. The following is one such refinement, by Kontorovich [2014].

Definition 4.1 ([Kontorovich, 2014]). Let $\mathcal{D}$ be a distribution over a domain X, and let $\rho : X^2 \to \mathbb{R}^{\geq 0}$. The *symmetrized distance* of $(X, \rho, \mathcal{D})$ is the random variable $\Xi = \xi \cdot \rho(x, x')$ where $x, x' \sim \mathcal{D}$ are independent and ξ is uniform on $\{\pm 1\}$ independent of x, x'. The *subgaussian diameter* of $(X, \rho, \mathcal{D})$, denoted $\Delta_{\mathrm{SG}}(X, \rho, \mathcal{D})$, is the smallest $\sigma \geq 0$ such that

$$
\mathbb{E} \left[e^{\lambda \Xi} \right] \leq e^{\sigma^2 \lambda^2 / 2}, \quad \forall \lambda \in \mathbb{R}.
$$

Kontorovich [2014] showed the following theorem:

Theorem 4.2 ([Kontorovich, 2014], informal). *Let $f : X^n \to \mathbb{R}$ be a function mapping databases of size n over a domain X to the reals. Assume that there exists a function $\rho : X^2 \to \mathbb{R}^{\geq 0}$ s.t. for every $i \in [n]$, every $S \in X^n$, and every $y, z \in X$ we have that*

$$
\left| f \left(S^{(i \leftarrow y)} \right) - f \left(S^{(i \leftarrow z)} \right) \right| \leq \rho(y, z),
$$

where $S^{(i \leftarrow x)}$ is the same as S except that the ith element is replaced with x. Then,

$$
\Pr_{S \sim \mathcal{D}^n} [|f(S) - f(\mathcal{D}^n)| \geq t] \leq 2 \exp \left(-\frac{t^2}{2n \cdot \Delta_{\mathrm{SG}}^2(X, \rho, \mathcal{D})} \right).
$$

Informally, using the above theorem it is possible to obtain concentration bounds for functions with unbounded sensitivity (in worst case), provided that the sensitivity (as a random variable) is subgaussian. In this section we show that our result implies a similar version of this theorem. While the bound we obtain is weaker than Theorem 4.2, our techniques can be extended to obtain concentration bounds even in cases where the sensitivity is *not* subgaussian (that is, in cases where the subgaussian diameter is unbounded, and hence, Theorem 4.2 could not be applied).

Let us denote $\sigma = \Delta_{\mathrm{SG}}(X, \rho, \mathcal{D})$. Now for $t \geq 0$,

$$\Pr_{x,y\sim\mathcal{D}}[\rho(x,y) \geq t] \leq 2 \Pr_{\substack{x,y\in\mathcal{D} \\ \xi\in\{\pm1\}}}[\xi \cdot \rho(x,y) \geq t] = 2\Pr[\Xi \geq t] = 2\Pr[e^{\frac{t}{\sigma^2}\cdot\Xi} \geq e^{\frac{t}{\sigma^2}\cdot t}]$$

$$\leq 2e^{-\frac{t^2}{\sigma^2}} \cdot \mathbb{E}\left[e^{\frac{t}{\sigma^2}\cdot\Xi}\right] \leq 2e^{-\frac{t^2}{\sigma^2}} \cdot e^{\frac{\sigma^2}{2}\cdot\frac{t^2}{\sigma^4}} = 2\exp\left(-\frac{t^2}{2\sigma^2}\right). \qquad (4.1)$$

So,

$$\mathbb{E}_{\substack{S\sim\mathcal{D}^n \\ x'\sim\mathcal{D}}}\left[\mathbb{1}\left\{\left|f(S) - f\left(S^{(i\leftarrow x')}\right)\right| > \lambda\right\} \cdot \left|f(S) - f\left(S^{(i\leftarrow x')}\right)\right|\right]$$

$$\leq \mathbb{E}_{x,y\sim\mathcal{D}}\left[\mathbb{1}\left\{\rho(x,y) > \lambda\right\} \cdot \rho(x,y)\right]$$

$$= \int_0^\lambda \Pr_{x,y\sim\mathcal{D}}\left[\mathbb{1}\left\{\rho(x,y) > \lambda\right\} \cdot \rho(x,y) \geq t\right]\mathrm{d}t \;+\; \int_\lambda^\infty \Pr_{x,y\sim\mathcal{D}}\left[\mathbb{1}\left\{\rho(x,y) > \lambda\right\} \cdot \rho(x,y) \geq t\right]\mathrm{d}t$$

$$= \int_0^\lambda \Pr_{x,y\sim\mathcal{D}}\left[\rho(x,y) \geq \lambda\right]\mathrm{d}t \;+\; \int_\lambda^\infty \Pr_{x,y\sim\mathcal{D}}\left[\rho(x,y) \geq t\right]\mathrm{d}t$$

$$= \lambda \cdot \Pr_{x,y\sim\mathcal{D}}\left[\rho(x,y) \geq \lambda\right] \;+\; \int_\lambda^\infty \Pr_{x,y\sim\mathcal{D}}\left[\rho(x,y) \geq t\right]\mathrm{d}t$$

$$\leq \lambda \cdot 2\exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \;+\; \int_\lambda^\infty 2\exp\left(-\frac{t^2}{2\sigma^2}\right)\mathrm{d}t = \lambda \cdot 2\exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \;+\; \sqrt{2\pi}\sigma \cdot \mathrm{erfc}\left(\frac{\lambda}{\sqrt{2}\sigma}\right)$$

$$\leq \lambda \cdot 2\exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \;+\; \sqrt{2\pi}\sigma \cdot \exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \leq 3(\lambda + \sigma) \cdot \exp\left(-\frac{\lambda^2}{2\sigma^2}\right) \triangleq \Delta.$$

In order to apply Theorem 1.2 we need to ensure that $n \geq O\left(\frac{1}{\varepsilon\cdot\min\{1,\varepsilon\}}\ln\left(\frac{\lambda\cdot\min\{1,\varepsilon\}}{\Delta}\right)\right)$. For our choice of Δ, it suffices to set $\varepsilon_0 = \Theta\left(\frac{\lambda}{\sqrt{n}\sigma}\right)$, assuming that $\frac{\lambda}{\sqrt{n}\sigma} \leq 1$. Otherwise, if $\frac{\lambda}{\sqrt{n}\sigma} > 1$, we will choose $\varepsilon_1 = \Theta\left(\frac{\lambda^2}{n\sigma^2}\right)$. Plugging (ε_0, Δ) or (ε_1, Δ) into Theorem 1.2, and simplifying, we get

$$\Pr_{S\sim\mathcal{D}}[|f(S) - f(\mathcal{D}^n)| \geq t] \leq \begin{cases} e^{-\Omega\left(\frac{t}{\sqrt{n}\sigma}\right)} & , \quad t \leq \sigma \cdot n^{1.5} \\ e^{-\Omega\left(\frac{t^{2/3}}{\sigma^{2/3}}\right)} & , \quad t > \sigma \cdot n^{1.5} \end{cases} \qquad (4.2)$$

Clearly, the bound of Theorem 4.2 is stronger. Note, however, that the only assumption we used here is that $\int_\lambda^\infty \Pr_{x,y\sim\mathcal{D}}[\rho(x,y) \geq t]\mathrm{d}t$ is small. Hence, as the following section shows, this argument could be extended to obtain concentration bounds even when $\Delta_{\mathrm{SG}}(X, \rho, \mathcal{D})$ is unbounded. We remark that Inequality 4.2 can be slightly improved by using part 2 of Theorem 1.2. This will be illustrated in the following section.

4.2. **Example: Concentration Under Infinite Variance.** Let $f : X^n \to \mathbb{R}$ be a function mapping databases of size n over a domain X to the reals. Assume that there exists a function $\rho : X^2 \to \mathbb{R}^{\geq 0}$ s.t. for every $i \in [n]$, every $S \in X^n$, and every $y, z \in X$ we have that

$$\left|f\left(S^{(i\leftarrow y)}\right) - f\left(S^{(i\leftarrow z)}\right)\right| \leq \rho(y, z),$$

where $S^{(i \leftarrow x)}$ is the same as S except that the ith element is replaced with x.

As stated in the previous section, the results of [Kontorovich, 2014] can be used to obtain a high probability bound on $|f(S) - f(\mathcal{D}^n)|$ whenever $\Pr_{x,y \sim \mathcal{D}}[\rho(x,y) \geq t] \leq \exp\left(-t^2/\sigma^2\right)$ for some $\sigma > 0$. In contrast, our bound can be used whenever $\int_\lambda^\infty \Pr_{x,y \sim \mathcal{D}}[\rho(x,y) \geq t] dt$ is finite. In particular, we now use it to obtain a concentration bound for a case where the probability distribution of $\rho(x,y)$ is heavy tailed, and in fact, has infinite variance. Specifically, assume that all we know on $\rho(x,y)$ is that $\Pr[\rho(x,y) \geq t] \leq 1/t^2$ for every $t \geq 1$ (this is a special case of the *Pareto distribution*, with infinite variance). Let $\lambda \geq 1$. We calculate:

$$\mathop{\mathbb{E}}_{\substack{S \sim \mathcal{D}^n \\ x' \sim \mathcal{D}}}\left[\mathbb{1}\left\{\left|f(S) - f\left(S^{(i \leftarrow x')}\right)\right| > \lambda\right\} \cdot \left|f(S) - f\left(S^{(i \leftarrow x')}\right)\right|\right]$$

$$\leq \mathop{\mathbb{E}}_{x,y \sim \mathcal{D}}\left[\mathbb{1}\left\{\rho(x,y) > \lambda\right\} \cdot \rho(x,y)\right]$$

$$= \int_0^\lambda \mathop{\Pr}_{x,y \sim \mathcal{D}}\left[\mathbb{1}\left\{\rho(x,y) > \lambda\right\} \cdot \rho(x,y) \geq t\right] dt \; + \; \int_\lambda^\infty \mathop{\Pr}_{x,y \sim \mathcal{D}}\left[\mathbb{1}\left\{\rho(x,y) > \lambda\right\} \cdot \rho(x,y) \geq t\right] dt$$

$$= \int_0^\lambda \mathop{\Pr}_{x,y \sim \mathcal{D}}\left[\rho(x,y) \geq \lambda\right] dt \; + \; \int_\lambda^\infty \mathop{\Pr}_{x,y \sim \mathcal{D}}\left[\rho(x,y) \geq t\right] dt$$

$$= \lambda \cdot \mathop{\Pr}_{x,y \sim \mathcal{D}}\left[\rho(x,y) \geq \lambda\right] \; + \; \int_\lambda^\infty \mathop{\Pr}_{x,y \sim \mathcal{D}}\left[\rho(x,y) \geq t\right] dt$$

$$\leq \lambda \frac{1}{\lambda^2} \; + \; \int_\lambda^\infty \frac{1}{t^2} dt = \frac{2}{\lambda} \triangleq \Delta.$$

In order to apply Theorem 1.2 we need to ensure that $n \geq O\left(\frac{1}{\varepsilon \cdot \min\{1,\varepsilon\}} \ln\left(\frac{\lambda \cdot \min\{1,\varepsilon\}}{\Delta} + 1\right)\right)$. Assuming that $n \geq \ln(\lambda)$, with our choice of Δ it suffices to set $\varepsilon = \Theta\left(\sqrt{\frac{1}{n} \ln(\lambda)}\right)$. Plugging ε and Δ into Theorem 1.2, and simplifying, we get

$$\mathop{\Pr}_{S \sim \mathcal{D}}\left[|f(S) - f(\mathcal{D}^n)| \geq t\right] \leq \tilde{O}\left(\frac{n^{3/2}}{t^2}\right). \tag{4.3}$$

Observe that the above bound decays as $1/t^2$. This should be contrasted with Markov's Inequality, which would decay as $1/t$. Recall the assumption that the variance of $\rho(x,y)$ is unbounded. Hence, the variance of $f(S)$ can also be unbounded, and Chebyshev's inequality could not be applied.

As we now explain, Inequality 4.3 can be improved using part 2 of Theorem 1.2. To that end, for a fixed database $S \in X^n$, we calculate:

$$\mathop{\mathbb{E}}_{y,z \sim \mathcal{D}}\left[\mathbb{1}\left\{\left|f(S^{(i \leftarrow y)}) - f\left(S^{(i \leftarrow z)}\right)\right| \leq \lambda\right\} \cdot \left|f(S^{(i \leftarrow y)}) - f\left(S^{(i \leftarrow z)}\right)\right|\right]$$

$$\leq \mathop{\mathbb{E}}_{y,z \sim \mathcal{D}}\left[\rho(y,z)\right] \leq \int_0^1 1 dt + \int_1^\infty \frac{1}{t^2} dt = 2 \triangleq \tau.$$

In order to apply part 2 of Theorem 1.2 we need to ensure that $n \geq O\left(\frac{\lambda}{\varepsilon \cdot \min\{1,\varepsilon\}\tau} \ln\left(\frac{\varepsilon\tau}{\Delta}\right)\right)$.

For our choice of Δ and τ, if $n \geq \lambda \ln(\lambda)$ then it suffices to set $\varepsilon_0 = \Theta\left(\sqrt{\frac{\lambda}{n}\ln(\lambda)}\right)$.

Otherwise, if $n < \lambda \ln(\lambda)$ then it suffices to set $\varepsilon_1 = \Theta\left(\frac{\lambda}{n}\ln(\lambda)\right)$. Plugging (ε_0, Δ) or (ε_1, Δ) into Theorem 1.2, and simplifying, we get

$$\Pr_{S \sim \mathcal{D}}\left[|f(S) - f(\mathcal{D}^n)| \geq t\right] \leq \begin{cases} \tilde{O}\left(\frac{n^2}{t^3}\right) & , \quad t \leq n \\ \tilde{O}\left(\frac{n}{t^2}\right) & , \quad t > n \end{cases}$$

4.3. **Example: Triangles in Random Graphs.** A random graph $G(N, p)$ on N vertices $1, 2, \ldots, N$ is defined by drawing an edge between each pair $1 \leq i < j \leq N$ independently with probability p. There are $n = \binom{N}{2}$ i.i.d. random variables $x_{\{i,j\}}$ representing the choices: $x_{\{i,j\}} = x_{\{j,i\}} = 1$ if the edge $\{i, j\}$ is drawn, and 0 otherwise. We will use $\mathcal{D}$ to denote the probability $\Pr_{x \sim \mathcal{D}}[x = 1] = p$ and $\Pr_{x \sim \mathcal{D}}[x = 0] = 1 - p$, and let $S = \left(x_{\{1,2\}}, \ldots, x_{\{n-1,n\}}\right) \sim \mathcal{D}^n$.

We say that three vertices i, j, ℓ form a triangle if there is an edge between any pair of them. Denote $f_{K_3}(S)$ the number of triangles in the graph defined by S. For a small constant α, we would like to have an exponential bound on the following probability

$$\Pr\left[f_{K_3}(S) \geq (1 + \alpha) \cdot f_{K_3}(\mathcal{D}^n)\right].$$

Specifically, we are interested in small values of $p = o(1)$ such that $f_{K_3}(\mathcal{D}^n) = \binom{N}{3}p^3 = \Theta\left(N^3 p^3\right) = o(N)$. The difficulty with this choice of p is that (in worst-case) adding a single edge to the graph can increase the number of triangles by $(N - 2)$, which is much larger then the expected number of triangles. Indeed, until the breakthrough work of Vu [2002], no general exponential bounds were known. After the work of Vu, the following sharp bound was presented by Kim and Vu [2004]:

Theorem 4.3 ([Kim and Vu, 2004], informal)**.** *Let α be a small constant. It holds that*

$$\exp\left(-\Theta\left(p^2 N^2 \log(1/p)\right)\right) \leq \Pr_{S \sim \mathcal{D}^n}\left[f_{K_3}(S) \geq (1 + \alpha) \cdot f_{K_3}(\mathcal{D}^n)\right] \leq \exp\left(-\Theta\left(p^2 N^2\right)\right).$$

In this section we show that our result can be used to analyze this problem. While the bound we obtain is weaker than Theorem 4.3, we find it interesting that the same technique from the last sections can also be applied here. To make things more concrete, we fix

$$p = N^{-3/4}.$$

In order to use our concentration bound, we start by analyzing the expected difference incurred to f_{K_3} by resampling a single edge. We will denote $\blacktriangle_{i,j}(S)$ as the number of triangles that are created (or deleted) by adding (or removing) the edge $\{i, j\}$. That is,

$$\blacktriangle_{i,j}(S) = \left|\left\{\ell \neq i, j \ : \ x_{\{i,\ell\}} = 1 \text{ and } x_{\{\ell,j\}} = 1\right\}\right|.$$

Observe that $\blacktriangle_{i,j}(S)$ does not depend on $x_{\{i,j\}}$. Moreover, observe that for every fixture of $i < j$ we have that $\blacktriangle_{i,j}(S)$ is the sum of $(N - 2)$ i.i.d. indicators, each equal to 1 with probability p^2.

Fix $S = \left(x_{\{1,2\}}, \ldots, x_{\{n-1,n\}}\right) \in \{0, 1\}^n$ and $x' \in \{0, 1\}$. We have that

$$\left|f_{K_3}(S) - f_{K_3}\left(S^{(\{i,j\} \leftarrow x')}\right)\right| = \begin{cases} 0 & , \quad x_{\{i,j\}} = x' \\ \blacktriangle_{i,j}(S) & , \quad x_{\{i,j\}} \neq x' \end{cases}$$

where $S^{(\{i,j\}\leftarrow x')}$ is the same as S except with $x_{\{i,j\}}$ replaced with x'. Fix $i < j$. We can now calculate

$$
\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n \\ x'\sim\mathcal{D}}}\left[\mathbb{1}\left\{\left|f_{K_3}(S) - f_{K_3}\left(S^{(\{i,j\}\leftarrow x')}\right)\right| > \lambda\right\}\cdot\left|f_{K_3}(S) - f_{K_3}\left(S^{(\{i,j\}\leftarrow x')}\right)\right|\right]
$$

$$
= \mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n \\ x'\sim\mathcal{D}}}\left[\mathbb{1}\left\{x_{\{i,j\}}\neq x'\right\}\cdot\mathbb{1}\left\{\blacktriangle_{i,j}(S) > \lambda\right\}\cdot\blacktriangle_{i,j}(S)\right]
$$

$$
= \mathop{\Pr}_{x_{\{i,j\}},x'\sim\mathcal{D}}\left[x_{\{i,j\}}\neq x'\right]\cdot\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\left[\mathbb{1}\left\{\blacktriangle_{i,j}(S) > \lambda\right\}\cdot\blacktriangle_{i,j}(S)\right]
$$

$$
= 2p(1-p)\cdot\left(\lambda\cdot\mathop{\Pr}_{S\sim\mathcal{D}^n}[\blacktriangle_{i,j}(S)\geq\lambda] + \int_\lambda^N\mathop{\Pr}_{S\sim\mathcal{D}^n}[\blacktriangle_{i,j}(S)\geq t]\mathrm{d}t\right)
$$

$$
\leq 2pN\cdot\mathop{\Pr}_{S\sim\mathcal{D}^n}[\blacktriangle_{i,j}(S)\geq\lambda]. \tag{4.4}
$$

Recall that $\blacktriangle_{i,j}(S)$ is the sum of $(N-2)$ i.i.d. indicators, each equal to 1 with probability p^2. We can upper bound the probability that $\blacktriangle_{i,j}(S)\geq\lambda$ with the probability that a sum of N such random variables is at least λ. We will use the following variant of the Chernoff bound, known as the Chernoff-Hoeffding theorem:

Theorem 4.4 ([Hoeffding, 1963]). *Let $X_1,\ldots,X_n$ be independent random variables where $\Pr[X_i = 1] = p$ and $\Pr[X_i = 0] = 1 - p$ for some $0 < p < 1$. Let k be s.t. $p < \frac{k}{n} < 1$. Then,*

$$
Pr\left[\sum_{i=1}^n X_i \geq k\right] \leq \exp\left(-n\cdot D\left(\frac{k}{n}\middle\|p\right)\right),
$$

where $D(a\|p)$ is the relative entropy between an a-coin and a p-coin (i.e. between the Bernoulli(a) and Bernoulli(p) distribution):

$$
D(a\|p) = a\cdot\log\left(\frac{a}{p}\right) + (1-a)\cdot\log\left(\frac{1-a}{1-p}\right).
$$

Using the Chernoff-Hoeffding theorem, for $p^2 N < \lambda < N$, we have

$$
(4.4) \leq 2pN\cdot\exp\left(-N\cdot D\left(\frac{\lambda}{N}\middle\|p^2\right)\right). \tag{4.5}
$$

Recall that we fixed $p = N^{-3/4}$, and let us denote $\lambda = N^\gamma$ for some constant $0 < \gamma < 1$. We get:

$$
(4.5) = 2pN\cdot\exp\left(-N\cdot D\left(N^{\gamma-1}\middle\|N^{-6/4}\right)\right). \tag{4.6}
$$

We will use the following claim:

Claim 4.5. *Fix constants $c > b > 0$. For $N \geq \max\{2^{1/b}, 2^{8/(c-b)}\}$ we have that*

$$
D\left(N^{-b}\middle\|N^{-c}\right) \geq \frac{c-b}{2}\cdot N^{-b}\cdot\log(N).
$$

Using Claim 4.5, for large enough N, we have that

$$(4.6) \leq \exp\left(-\frac{1}{8} \cdot N^\gamma \log(N)\right). \tag{4.7}$$

So, denoting $\Delta = \exp\left(-\frac{1}{8} \cdot N^\gamma \log(N)\right)$, we get that

$$\mathop{\mathbb{E}}_{\substack{S \sim \mathcal{D}^n \\ x' \sim \mathcal{D}}}\left[\mathbb{1}\left\{\left|f_{K_3}(S) - f_{K_3}\left(S^{(\{i,j\}\leftarrow x')}\right)\right| > \lambda\right\} \cdot \left|f_{K_3}(S) - f_{K_3}\left(S^{(\{i,j\}\leftarrow x')}\right)\right|\right] \leq \Delta.$$

In order to obtain a meaningful bound, we will need to use part 2 of Theorem 1.2. To that end, for every fixture of $S \in X^n$ and $i < j$ we can compute

$$\mathop{\mathbb{E}}_{y,z \sim \mathcal{D}}\left[\mathbb{1}\left\{\left|f_{K_3}(S^{(\{i,j\}\leftarrow y)}) - f_{K_3}\left(S^{(\{i,j\}\leftarrow z)}\right)\right| \leq \lambda\right\} \cdot \left|f_{K_3}(S^{(\{i,j\}\leftarrow y)}) - f_{K_3}\left(S^{(\{i,j\}\leftarrow z)}\right)\right|\right]$$

$$\leq \mathop{\mathbb{E}}_{y,z \sim \mathcal{D}}\left[\mathbb{1}\left\{y \neq z\right\} \cdot \lambda\right] = 2p(1-p)\lambda \leq 2p\lambda \triangleq \tau.$$

Finally, we need to ensure that $n \geq O\left(\frac{\lambda}{\varepsilon \min\{1,\varepsilon\}\tau} \ln\left(\frac{\min\{1,\varepsilon\}\tau}{\Delta}\right)\right)$ in order to apply Theorem 1.2. With our choices for Δ and τ, it suffices to set $\varepsilon = \Theta\left(\sqrt{\frac{\lambda}{np}}\right)$. Plugging ε, Δ and τ into Theorem 1.2, and simplifying, we get that there exists a constant $c > 0$ s.t.

$$\mathop{\Pr}_{S \sim \mathcal{D}^n}\left[|f_{K_3}(S) - f_{K_3}(\mathcal{D}^n)| \geq c \cdot N^{5/8 + 3\gamma/2}\right] < \exp\left(-N^\gamma\right). \tag{4.8}$$

Recall that for our choice of $p = N^{-3/4}$ we have that $f_{K_3}(\mathcal{D}^n) = \Theta(N^{3/4})$, and hence, Inequality (4.8) gives an exponential bound on the probability that the number of triangles deviates from its expectation. For example, setting $\gamma = 1/13$ we have

$$\mathop{\Pr}_{S \sim \mathcal{D}^n}\left[|f_{K_3}(S) - f_{K_3}(\mathcal{D}^n)| \geq o\left(f_{K_3}(\mathcal{D}^n)\right)\right] < \exp\left(-N^{1/13}\right).$$

It remains to prove Claim 4.5:

Claim 4.5. *Fix constants $c > b > 0$. For $N \geq \max\{2^{1/b}, 2^{8/(c-b)}\}$ we have that*

$$D\left(N^{-b} \,\middle\|\, N^{-c}\right) \geq \frac{c-b}{2} \cdot N^{-b} \cdot \log(N).$$

Proof of Claim 4.5.

$$D\left(N^{-b} \,\middle\|\, N^{-c}\right) = N^{-b} \cdot \log\left(N^{c-b}\right) + \left(1 - N^{-b}\right) \cdot \log\left(\frac{1 - N^{-b}}{1 - N^{-c}}\right)$$

$$= N^{-b} \cdot \log\left(N^{c-b}\right) + \left(1 - N^{-b}\right) \cdot \log\left(\frac{N^c - N^{c-b}}{N^c - 1}\right)$$

$$= N^{-b} \cdot \log\left(N^{c-b}\right) + \left(1 - N^{-b}\right) \cdot \log\left(1 - \frac{N^{c-b} - 1}{N^c - 1}\right) \tag{4.9}$$

Using the fact that $\log(1 - x) \geq -2x$ for every $0 \leq x \leq \frac{1}{2}$, and assuming that $N \geq 2^{1/b}$, we have that

$$(4.9) \geq N^{-b} \cdot \log\left(N^{c-b}\right) - 2\left(1 - N^{-b}\right) \cdot \frac{N^{c-b} - 1}{N^c - 1}$$

$$= N^{-b} \cdot \log\left(N^{c-b}\right) - 2 \cdot \frac{N^{c-b} - 1}{N^c - 1} + 2N^{-b} \cdot \frac{N^{c-b} - 1}{N^c - 1}$$

$$\geq N^{-b} \cdot \log\left(N^{c-b}\right) - 2 \cdot \frac{N^{c-b} - 1}{N^c - 1}$$

$$\geq N^{-b} \cdot \log\left(N^{c-b}\right) - 2 \cdot \frac{N^{c-b}}{\frac{1}{2}N^c}$$

$$\geq N^{-b} \cdot \log\left(N^{c-b}\right) - 4N^{-b} \tag{4.10}$$

Assuming that $N \geq 2^{8/(c-b)}$ we get

$$(4.10) \geq \frac{1}{2} \cdot N^{-b} \cdot \log\left(N^{c-b}\right)$$

$$\geq \frac{c-b}{2} \cdot N^{-b} \cdot \log\left(N\right).$$

$$\square$$

5. Privately Identifying a High-Sensitivity Function

In this section we construct an example for a differentially private algorithm that, given a sample S, identifies a function f such that $|f(S)|$ is arbitrarily large, even though f is highly concentrated around zero (for a fresh sample from the underlying distribution). We interpret this result as an indication that differential privacy can be used to prove concentration bounds even for functions for which it does not guarantee generalization under adaptive selection.

Theorem 5.1. *Fix $\beta, \varepsilon, B > 0$, let $\mathcal{U}$ be the uniform distribution over $X = \{\pm 1\}^d$ where $d = \mathrm{poly}(1/\beta)$, and let $n \geq O(\frac{1}{\varepsilon^2} \ln(1/\beta))$. There exists an $(\varepsilon, 0)$-differentially private algorithm $\mathcal{A}$ that operates on a database $S \in (\{\pm 1\}^d)^n$ and returns a function mapping $(\{\pm 1\}^d)^n$ to $\mathbb{R}$, s.t. the following hold.*
(1) For every f in the range of $\mathcal{A}$ it holds that $\Pr_{S' \sim \mathcal{U}^n}[f(S') \neq f(\mathcal{U}^n)] \leq \beta$.
(2) $\Pr_{\substack{S \sim \mathcal{U}^n \\ f \leftarrow \mathcal{A}(S)}} [|f(S) - f(\mathcal{U}^n)| \geq B] \geq 1/2$.

Proof. For $t \in [d]$, define $f_t : (\{\pm 1\}^d)^n \to \mathbb{R}$ as

$$f_t(x_1, \ldots, x_n) = \begin{cases} 0 & , & \left|\sum_{i \in [n]} x_{i,t}\right| \leq \sqrt{2n \ln(2/\beta)} \\ B & , & \sum_{i \in [n]} x_{i,t} > \sqrt{2n \ln(2/\beta)} \\ -B & , & \sum_{i \in [n]} x_{i,t} < -\sqrt{2n \ln(2/\beta)} \end{cases}$$

That is, given a database S of n rows from $\{\pm 1\}^d$, we define $f_t(S)$ as 0 if the sum of column t (in absolute value) is less than some threshold, and otherwise set $f_t(S)$ to be $\pm B$

(depending on the sign of the sum). Observe that the global sensitivity of f_t is B, and that $f_t(\mathcal{U}^n) \triangleq \mathop{\mathbb{E}}_{S' \sim \mathcal{U}^n} [f_t(S')] = 0$. Also, by the Hoeffding bound, we have that

$$\Pr_{S \sim \mathcal{U}^n} [f_t(S) \neq 0] \leq \beta.$$

So, for every fixed t, with high probability over sampling $S \sim \mathcal{U}^n$ we have that $f_t(S) = 0 = f_t(\mathcal{U}^n)$. Nevertheless, as we now explain, if d is large enough, then an $(\varepsilon, 0)$-differentially private algorithm can easily identify a "bad" index t^* such that $|f_{t^*}(S)| = B$.

Consider the algorithm that on input $S = (x_1, x_2, \ldots, x_n)$ samples an index $t \in [d]$ with probability proportional to $\exp\left(\frac{\varepsilon}{4} \left|\sum_{i \in [n]} x_{i,t}\right|\right)$. We will call it algorithm `BadIndex`.

By the properties of the exponential mechanism, algorithm `BadIndex` is $(\varepsilon, 0)$-differentially private. Moreover, with probability at least $3/4$, the output t^* satisfies

$$\left|\sum_{i \in [n]} x_{i,t^*}\right| \geq \max_{t \in [d]} \left\{\left|\sum_{i \in [n]} x_{i,t}\right|\right\} - \frac{4}{\varepsilon} \ln(4d). \tag{5.1}$$

In addition, by Theorem 2.6 (tightness of Chernoff bound), for every fixed t it holds that

$$\Pr\left[\sum_{i \in [n]} x_{i,t} \geq 1.11 \cdot \sqrt{2n \ln(2/\beta)}\right] \geq \left(\frac{\beta}{2}\right)^{45}.$$

As the columns are independent, taking $d = 2\left(\frac{2}{\beta}\right)^{45}$, we get that

$$\Pr\left[\max_{t \in [d]} \left\{\sum_{i \in [n]} x_{i,t}\right\} \geq 1.11 \cdot \sqrt{2n \ln(2/\beta)}\right] \geq 3/4. \tag{5.2}$$

Combining (5.1) and (5.2) we get that with probability at least $1/2$ algorithm `BadIndex` identifies an index t^* such that

$$\left|\sum_{i \in [n]} x_{i,t^*}\right| \geq 1.11 \cdot \sqrt{2n \ln(2/\beta)} - \frac{4}{\varepsilon} \ln(4d).$$

Assuming that $n \geq O(\frac{1}{\varepsilon^2} \ln(1/\beta))$ we get that with probability at least $1/2$ algorithm `BadIndex` outputs an index t^* s.t. $f_{t^*}(S) = B$. □

5.1. Max-Information.

In this section we show that algorithm `BadIndex` has relatively high *max-information*: Given two (correlated) random variables Y, Z, we use $Y \otimes Z$ denote the random variable obtained by drawing independent copies of Y and Z from their respective marginal distributions.

Definition 5.2 (Max-Information [Dwork et al., 2015a])**.** Let Y and Z be jointly distributed random variables over the domain $(\mathcal{Y}, \mathcal{Z})$. The β-approximate max-information between Y and Z is defined as

$$I_\infty^\beta(Y; Z) = \log \sup_{\substack{\mathcal{O} \subseteq (\mathcal{Y} \times \mathcal{Z}), \\ \Pr[(Y,Z) \in \mathcal{O}] > \beta}} \frac{\Pr[(Y, Z) \in \mathcal{O}] - \beta}{\Pr[Y \otimes Z \in \mathcal{O}]}.$$

An algorithm $\mathcal{A} : X^n \to F$ has β-approximate max-information of k over product distributions, written $I_{\infty,P}^{\beta}(\mathcal{A}, n) \leq k$, if for every distribution $\mathcal{D}$ over X, we have $I_{\infty}^{\beta}(S; \mathcal{A}(S)) \leq k$ when $S \sim \mathcal{D}^n$.

It follows immediately from the definition that approximate max-information controls the probability of "bad events" that can happen as a result of the dependence of $\mathcal{A}(S)$ on S: for every event $\mathcal{O}$, we have $\Pr[(S, \mathcal{A}(S)) \in \mathcal{O}] \leq 2^k \Pr[S \otimes \mathcal{A}(S) \in \mathcal{O}] + \beta$.

Consider again algorithm $\texttt{BadIndex} : (\{\pm 1\})^n \to F$ that operates on database S of size $n = O(\frac{1}{\varepsilon^2} \ln(1/\beta))$ and identifies, with probability $1/2$, a function f s.t. $f(S) \neq 0$, even though $f(S') = 0$ w.p. $1 - \beta$ for a fresh sample S'. Let us define $\mathcal{O}$ as the set of all pairs (S, f), where S is a database and f is a function in the range of algorithm $\texttt{BadIndex}$ such that $f(S) \neq 0$. That is,

$$\mathcal{O} = \{(S, f) \in (\{\pm 1\})^n \times F \; : \; f(S) \neq 0\}.$$

If we assume that $I_{\infty,P}^{1/4}(\texttt{BadIndex}, n) \leq k$, then by Definition 5.2 we have:

$$\frac{1}{2} \leq \Pr_{\substack{S \sim \mathcal{U}^n \\ f \leftarrow \texttt{BadIndex}(S)}} [(S, f) \in \mathcal{O}] \leq e^k \cdot \Pr_{\substack{S \sim \mathcal{U}^n \\ T \sim \mathcal{U}^n \\ f \leftarrow \texttt{BadIndex}(T)}} [(S, f) \in \mathcal{O}] + \frac{1}{4} \leq e^k \cdot \beta + \frac{1}{4}.$$

So $k \geq \ln(\frac{1}{4\beta}) = \Omega(\varepsilon^2 n)$.

References

R. Bassily, K. Nissim, A. D. Smith, T. Steinke, U. Stemmer, and J. Ullman. Algorithmic stability for adaptive data analysis. In D. Wichs and Y. Mansour, editors, *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 1046–1059. ACM, 2016. ISBN 978-1-4503-4132-5. doi: 10.1145/2897518.2897566. URL http://doi.acm.org/10.1145/2897518.2897566.

H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, 23:493–507, 1952.

C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In S. Vaudenay, editor, *EUROCRYPT*, volume 4004 of *Lecture Notes in Computer Science*, pages 486–503. Springer, 2006a.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006b.

C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth. Generalization in adaptive data analysis and holdout reuse. In *Advances in Neural Information Processing Systems (NIPS)*, Montreal, December 2015a.

C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth. Preserving statistical validity in adaptive data analysis. In *ACM Symposium on the Theory of Computing (STOC)*. ACM, June 2015b.

W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

J. H. Kim and V. H. Vu. Divide and conquer martingales and the number of triangles in a random graph. *Random Structures and Algorithms*, 24(2):166–174, 2004. ISSN 1098-2418. doi: 10.1002/rsa.10113. URL http://dx.doi.org/10.1002/rsa.10113.

P. N. Klein and N. E. Young. On the number of iterations for dantzig-wolfe optimization and packing-covering approximation algorithms. *SIAM J. Comput.*, 44(4):1154–1172, 2015. doi: 10.1137/12087222X. URL http://dx.doi.org/10.1137/12087222X.

A. Kontorovich. Concentration in unbounded metric spaces and algorithmic stability. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, volume 32 of *JMLR Workshop and Conference Proceedings*, pages 28–36. JMLR.org, 2014. URL http://jmlr.org/proceedings/papers/v32/kontorovicha14.html.

F. McSherry and K. Talwar. Mechanism design via differential privacy. In *FOCS*, pages 94–103. IEEE, Oct 20–23 2007.

T. Steinke and J. Ullman. Subgaussian tail bounds via stability arguments. *ArXiv.org*, (arXiv:1701.03493 [cs.DM]), 2017.

V. H. Vu. Concentration of non-lipschitz functions and applications. *Random Structures and Algorithms*, 20(3):262–316, 2002. doi: 10.1002/rsa.10032. URL http://dx.doi.org/10.1002/rsa.10032.

Appendix A. Missing Details

Claim 3.3. *Fix a function $f : X^n \to \mathbb{R}$ and parameters $\varepsilon, \lambda \geq 0$. If $M : (X^n)^T \to Y$ is $(\varepsilon, (f, \lambda))$-differentially private then for every (f, λ)-neighboring databases $\vec{S}, \vec{S'} \in (X^n)^T$ and every function $h : Y \to \mathbb{R}$ we have that*

$$\underset{y \leftarrow M(\vec{S})}{\mathbb{E}} [h(y)] \leq e^{-\varepsilon} \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ (e^{\varepsilon} - e^{-\varepsilon}) \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [|h(y)|].$$

Proof.

$$\underset{y \leftarrow M(\vec{S})}{\mathbb{E}} [h(y)] = \int_0^\infty \underset{y \leftarrow M(\vec{S})}{\Pr} [h(y) \geq z] \mathrm{d}z \ - \ \int_{-\infty}^0 \underset{y \leftarrow M(\vec{S})}{\Pr} [h(y) \leq z] \mathrm{d}z$$

$$\leq e^{\varepsilon} \cdot \int_0^\infty \underset{y \leftarrow M(\vec{S'})}{\Pr} [h(y) \geq z] \mathrm{d}z \ - \ e^{-\varepsilon} \cdot \int_{-\infty}^0 \underset{y \leftarrow M(\vec{S'})}{\Pr} [h(y) \leq z] \mathrm{d}z$$

$$= e^{-\varepsilon} \left[\int_0^\infty \underset{y \leftarrow M(\vec{S'})}{\Pr} [h(y) \geq z] \mathrm{d}z \ - \ \int_{-\infty}^0 \underset{y \leftarrow M(\vec{S'})}{\Pr} [h(y) \leq z] \mathrm{d}z \right]$$

$$+ (e^{\varepsilon} - e^{-\varepsilon}) \cdot \int_0^\infty \underset{y \leftarrow M(\vec{S'})}{\Pr} [h(y) \geq z] \mathrm{d}z$$

$$= e^{-\varepsilon} \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ (e^{\varepsilon} - e^{-\varepsilon}) \cdot \int_0^\infty \underset{y \leftarrow M(\vec{S'})}{\Pr} [h(y) \geq z] \mathrm{d}z$$

$$\leq e^{-\varepsilon} \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ (e^{\varepsilon} - e^{-\varepsilon}) \cdot \int_0^\infty \underset{y \leftarrow M(\vec{S'})}{\Pr} [|h(y)| \geq z] \mathrm{d}z$$

$$= e^{-\varepsilon} \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ (e^{\varepsilon} - e^{-\varepsilon}) \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [|h(y)|]$$

$\square$

Remark A.1. *To get the version of Claim 3.3 stated in the main body of the paper (that is, without the $e^{-\varepsilon}$ factor) observe that if $\underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \geq 0$ then $e^{-\varepsilon} \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \leq \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)]$. Otherwise, for every $0 \leq \varepsilon \leq 1$ we have*

$$\underset{y \leftarrow M(\vec{S})}{\mathbb{E}} [h(y)] \leq e^{-\varepsilon} \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ (e^{\varepsilon} - e^{-\varepsilon}) \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [|h(y)|]$$

$$\leq (1 - \varepsilon) \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ 3\varepsilon \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [|h(y)|] \leq \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [h(y)] \ + \ 4\varepsilon \cdot \underset{y \leftarrow M(\vec{S'})}{\mathbb{E}} [|h(y)|].$$

A.1. **Multi Sample Expectation Bound.**

Lemma A.1 (Expectation Bound). *Let $\mathcal{D}$ be a distribution over a domain X, let $f : X^n \to \mathbb{R}$, and let Δ, λ, τ be s.t.*

$$\forall i \in [n], \quad \underset{\substack{S \sim \mathcal{D}^n \\ z \sim \mathcal{D}}}{\mathbb{E}} \left[\mathbb{1}_{|f(S) - f(S^{(i \leftarrow z)})| > \lambda} \cdot \left| f(S) - f\left(S^{(i \leftarrow z)}\right) \right| \right] \leq \Delta,$$

and,

$$\forall i \in [n], \quad \underset{S_1, \ldots, S_T \sim \mathcal{D}^n}{\mathbb{E}} \underset{t \in [T]}{\max} \underset{y, z \sim \mathcal{D}}{\mathbb{E}} \left[\mathbb{1}_{\left| f(S_t^{(i \leftarrow y)}) - f\left(S_t^{(i \leftarrow z)}\right) \right| \leq \lambda} \cdot \left| f(S_t^{(i \leftarrow y)}) - f\left(S_t^{(i \leftarrow z)}\right) \right| \right] \leq \tau,$$

where $S^{(i \leftarrow z)}$ is the same as S except that the ith element is replaced with z. Let $\mathcal{A} : (X^n)^T \rightarrow ([T] \cup \perp)$ be an $(\varepsilon, (f, \lambda))$-differentially private algorithm that operates on T databases of size n from X, and outputs an index $1 \leq t \leq T$ or $\perp$. Then

$$\left| \underset{\substack{\vec{S} \sim \mathcal{D}^{nT} \\ t \leftarrow \mathcal{A}(\vec{S})}}{\mathbb{E}} [\mathbb{1}\{t \neq \perp\} \cdot (f(\mathcal{D}^n) - f(S_t))] \right| \leq (e^{\varepsilon} - e^{-\varepsilon}) \cdot \tau n + 6\Delta nT.$$

Proof. Let $\vec{S}' = (S'_1, \ldots, S'_T) \sim \mathcal{D}^{nT}$ be independent of $\vec{S}$. Recall that each element S_t of $\vec{S}$ is itself a vector $(x_{t,1}, \ldots, x_{t,n})$, and the same is true for each element S'_t of $\vec{S}'$. We will sometimes refer to the vectors $S_1, \ldots, S_T$ as the *subsamples* of $\vec{S}$.

We define a sequence of intermediate samples that allow us to interpolate between $\vec{S}$ and $\vec{S}'$. Formally, for $\ell \in \{0, 1, \ldots, n\}$ define $\vec{S}^\ell = (S_1^\ell, \ldots, S_T^\ell) \in (X^n)^T$ where $S_t^\ell = (x_{t,1}^\ell, \ldots, x_{t,n}^\ell)$ and

$$x_{t,i}^\ell = \begin{cases} x_{t,i} & , & i > \ell \\ x'_{t,i} & , & i \leq \ell \end{cases}$$

That is, every subsample S_t^ℓ of $\vec{S}^\ell$ is identical to S'_t on the first ℓ elements, and identical to S_t thereafter. By construction we have $\vec{S}^0 = \vec{S}$ and $\vec{S}^n = \vec{S}'$. Moreover, for every t we have that S_t^ℓ and $S_t^{\ell-1}$ differ in exactly one element. In terms of these intermediate samples we can write:

$$\left| \underset{\vec{S} \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{t \leftarrow \mathcal{A}(\vec{S})}{\mathbb{E}} [\mathbb{1}\{t \neq \perp\} \cdot (f(\mathcal{D}^n) - f(S_t))] \right|$$

$$= \left| \underset{\vec{S} \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{t \leftarrow \mathcal{A}(\vec{S})}{\mathbb{E}} \left[\mathbb{1}\{t \neq \perp\} \cdot \left(\underset{\vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} [f(S'_t)] - f(S_t) \right) \right] \right|$$

$$= \left| \underset{\vec{S} \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{t \leftarrow \mathcal{A}(\vec{S})}{\mathbb{E}} \underset{\vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \left[\mathbb{1}\{t \neq \perp\} \cdot (f(S'_t) - f(S_t)) \right] \right|$$

$$= \left| \sum_{\ell \in [n]} \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{t \leftarrow \mathcal{A}(\vec{S})}{\mathbb{E}} \left[\mathbb{1}\{t \neq \perp\} \cdot \left(f(S_t^\ell) - f(S_t^{\ell-1}) \right) \right] \right|$$

$$\leq \sum_{\ell \in [n]} \left| \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{t \leftarrow \mathcal{A}(\vec{S})}{\mathbb{E}} \left[\mathbb{1}\{t \neq \perp\} \cdot \left(f(S_t^\ell) - f(S_t^{\ell-1}) \right) \right] \right|$$

$$= \sum_{\ell \in [n]} \left| \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{Z \sim \mathcal{D}^T}{\mathbb{E}} \underset{t \leftarrow \mathcal{A}(\vec{S})}{\mathbb{E}} \left[\mathbb{1}\{t \neq \perp\} \cdot \left(f(S_t^\ell) - f(S_t^{\ell-1}) \right) \right] \right| \tag{A.1}$$

Given a multisample $\vec{S} = (S_1, \ldots, S_T) \in (X^n)^T$, a vector $Z = (z_1 \ldots, z_T) \in X^T$, and an index $1 \leq k \leq n$, we define $\vec{S}^{(k \leftarrow Z)}$ to be the same as $\vec{S}$ except that the kth element of *every* subsample S_i is replaced with z_i. Observe that by construction, for every ℓ, Z we have $\vec{S}^{\ell, (\ell \leftarrow Z)} = \vec{S}^{\ell-1, (\ell \leftarrow Z)}$. Thus,

$$(A.1) = \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S})} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f(S_t^\ell) - f\left(S_t^{\ell,(\ell \leftarrow Z)}\right) \right) \right. \right.$$

$$\left. \left. - \mathbb{1}\{t \neq \bot\} \cdot \left(f(S_t^{\ell-1}) - f\left(S_t^{\ell-1,(\ell \leftarrow Z)}\right) \right) \right] \right|. \quad (A.2)$$

Observe that the pairs $(\vec{S}, \vec{S}^\ell)$ and $\left(\vec{S}, \vec{S}^{\ell,(\ell \leftarrow Z)}\right)$ are identically distributed. Namely, both $\vec{S}^\ell$ and $\vec{S}^{\ell,(\ell \leftarrow Z)}$ agree with $\vec{S}$ on the last $(n - \ell)$ entries of every subsample, and otherwise contain i.i.d. samples from $\mathcal{D}$. Hence, the expectation of $\left(f(S_t^\ell) - f\left(S_t^{\ell,(\ell \leftarrow Z)}\right) \right)$ is zero, and we get

$$(A.2) = \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S})} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{\ell-1,(\ell \leftarrow Z)}\right) - f(S_t^{\ell-1}) \right) \right] \right|. \quad (A.3)$$

Observe that the pair $(\vec{S}^{\ell-1}, \vec{S})$ has the same distribution as $(\vec{S}, \vec{S}^{\ell-1})$. Specifically, the first component is nT independent samples from $\mathcal{D}$ and the second component is equal to the first component with a subset of the entries replaced by fresh independent samples from $\mathcal{D}$. Thus,

$$(A.3) = \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^{\ell-1})} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right|$$

$$\leq \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda \\ \text{and} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^{\ell-1})} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right|$$

$$+ \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \quad (A.4)$$

When $\max_{m\in[T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda$ we now use the properties of algorithm $\mathcal{A}$ to argue that $\mathcal{A}(\vec{S}^{\ell-1}) \approx \mathcal{A}(\vec{S}^\ell)$. Be Claim 3.3 we get that

$$(A.4) \leq \sum_{\ell\in[n]} \left| \mathbb{E}_{\vec{S},\vec{S}'\sim\mathcal{D}^{nT}} \mathbb{E}_{Z\sim\mathcal{D}^T} \left[\mathbb{1}\left\{ \begin{array}{c} \max_{m\in[T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda \\ \text{and} \\ \max_{m\in[T]} |f\left(S_m^{(\ell\leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \mathbb{E}_{t\leftarrow\mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell\leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right|$$

$$+ (e^\varepsilon - e^{-\varepsilon}) \cdot \sum_{\ell\in[n]} \left| \mathbb{E}_{\vec{S},\vec{S}'\sim\mathcal{D}^{nT}} \mathbb{E}_{Z\sim\mathcal{D}^T} \left[\mathbb{1}\left\{ \begin{array}{c} \max_{m\in[T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda \\ \text{and} \\ \max_{m\in[T]} |f\left(S_m^{(\ell\leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \mathbb{E}_{t\leftarrow\mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \bot\} \cdot \left| f\left(S_t^{(\ell\leftarrow Z)}\right) - f(S_t) \right| \right] \right] \right|$$

$$+ \sum_{\ell\in[n]} \left| \mathbb{E}_{\vec{S},\vec{S}'\sim\mathcal{D}^{nT}} \mathbb{E}_{Z\sim\mathcal{D}^T} \left[\mathbb{1}\left\{ \begin{array}{c} \max_{m\in[T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m\in[T]} |f\left(S_m^{(\ell\leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \max_{m\in[T]} \left| f\left(S_m^{(\ell\leftarrow Z)}\right) - f(S_m) \right| \right] \right| \tag{A.5}$$

We can remove one of the two requirements in the indicator function in the middle row (this makes the expression bigger), to get:

$$(A.5) \leq \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda \\ \text{and} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \perp\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right|$$

$$+ (e^\varepsilon - e^{-\varepsilon}) \cdot \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\left\{ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \right\} \cdot \right. \right.$$

$$\left. \left. \cdot \mathbb{1}\{t \neq \perp\} \cdot \left| f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right| \right] \right|$$

$$+ \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \tag{A.6}$$

Furthermore, we can replace $\mathbb{1} \left\{ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \right\}$ in the middle row with the weaker requirement – just for the specific t that was selected by algorithm $\mathcal{A}$. This yields:

$$(A.6) \leq \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^{\ell})| \leq \lambda \\ \text{and} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right|$$

$$+ (e^\varepsilon - e^{-\varepsilon}) \cdot \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1} \left\{ |f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t)| \leq \lambda \right\} \cdot \right. \right.$$

$$\left. \left. \cdot \mathbb{1}\{t \neq \bot\} \cdot \left| f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right| \right] \right|$$

$$+ \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^{\ell})| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \tag{A.7}$$

Using the fact that the pairs $(\vec{S}, \vec{S}^\ell)$ and $(\vec{S}^\ell, \vec{S})$ are identically distributed, we can switch them in the middle row, to get

$$(A.7) \leq \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda \\ \text{and} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right. \cdot$$

$$\left. \left. \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right|$$

$$+ (e^\varepsilon - e^{-\varepsilon}) \cdot \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S})} \mathop{\mathbb{E}}_{\substack{\vec{S}' \sim \mathcal{D}^{nT} \\ Z \sim \mathcal{D}^T}} \left[\mathbb{1}\left\{ |f\left(S_t^{\ell, (\ell \leftarrow Z)}\right) - f(S_t^\ell)| \leq \lambda \right\} \cdot \right. \right.$$

$$\left. \left. \cdot \mathbb{1}\{t \neq \bot\} \cdot \left| f\left(S_t^{\ell, (\ell \leftarrow Z)}\right) - f(S_t^\ell) \right| \right] \right|$$

$$+ \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right. \cdot$$

$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \tag{A.8}$$

Using our assumptions on the function f and the distribution $\mathcal{D}$, the middle row is at most $(e^\varepsilon - e^{-\varepsilon})n\tau$. This brings us to:

$$(A.8) \leq \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \leq \lambda \\ \text{and} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| \leq \lambda \end{array} \right\} \right. \right. $$

$$ \left. \left. \cdot \mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right| $$

$$+ (e^\varepsilon - e^{-\varepsilon})n\tau + \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right. $$

$$ \left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \tag{A.9}$$

Our next task is to remove the indicator function in the first row. This is useful as the pairs $\left(\vec{S}^\ell, \vec{S}^{(\ell \leftarrow Z)} \right)$ and $(\vec{S}^\ell, \vec{S})$ are identically distributed, and hence, if we were to remove the indicator function, the first row would be equal to zero. To that end we add and subtract the first row with the complementary indicator function (this amounts to multiplying the last summation by 2). We get

$$(A.9) \leq \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathop{\mathbb{E}}_{t \leftarrow \mathcal{A}(\vec{S}^\ell)} \left[\mathbb{1}\{t \neq \bot\} \cdot \left(f\left(S_t^{(\ell \leftarrow Z)}\right) - f(S_t) \right) \right] \right] \right| $$

$$+ (e^\varepsilon - e^{-\varepsilon})n\tau + 2 \cdot \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S}' \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right. $$

$$ \left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \tag{A.10}$$

Now the first row is 0, so

$$(A.10) = (e^\varepsilon - e^{-\varepsilon})n\tau + 2 \cdot \sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S'} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{or} \\ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \end{array} \right\} \right. \right.$$

$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right| \qquad (A.11)$$

We can replace the *or* condition in the indicator function with the sum of the two conditions:

$$(A.11) \leq (e^\varepsilon - e^{-\varepsilon})n\tau$$

$$+ 2\sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S'} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \right\} \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$+ 2\sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S'} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \max_{m \in [T]} |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \right\} \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$(A.12)$$

In the third row, we can replace $\max_{m \in [T]}$ with $\sum_{m \in [T]}$, to get

$$(A.12) \leq (e^\varepsilon - e^{-\varepsilon})n\tau$$

$$+ 2\sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S'} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \right\} \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$+ 2\sum_{\ell \in [n]} \sum_{m \in [T]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S'} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ |f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m)| > \lambda \right\} \cdot \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$(A.13)$$

Applying our assumptions on f and $\mathcal{D}$ to the third row brings us to

$$(A.13) \leq (e^\varepsilon - e^{-\varepsilon})n\tau + 2nT\Delta$$

$$+ 2\sum_{\ell \in [n]} \left| \mathop{\mathbb{E}}_{\vec{S},\vec{S'} \sim \mathcal{D}^{nT}} \mathop{\mathbb{E}}_{Z \sim \mathcal{D}^T} \left[\mathbb{1} \left\{ \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \right\} \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$(A.14)$$

The issue now is that the expression inside the indicator function is different from the expression outside of it. To that end, we split the indicator function as follows:

$$(A.14) \leq (e^\varepsilon - e^{-\varepsilon})n\tau + 2nT\Delta$$

$$+ 2\sum_{\ell \in [n]} \left| \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{Z \sim \mathcal{D}^T}{\mathbb{E}} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{and} \\ \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| > \lambda \end{array} \right\} \cdot \right. \right.$$
$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$+ 2 \cdot \sum_{\ell \in [n]} \left| \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{Z \sim \mathcal{D}^T}{\mathbb{E}} \left[\mathbb{1} \left\{ \begin{array}{c} \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \\ \text{and} \\ \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \leq \lambda \end{array} \right\} \cdot \right. \right.$$
$$\left. \left. \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$\leq (e^\varepsilon - e^{-\varepsilon})n\tau + 2nT\Delta$$

$$+ 2\sum_{\ell \in [n]} \left| \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{Z \sim \mathcal{D}^T}{\mathbb{E}} \left[\mathbb{1} \left\{ \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| > \lambda \right\} \cdot \max_{m \in [T]} \left| f\left(S_m^{(\ell \leftarrow Z)}\right) - f(S_m) \right| \right] \right|$$

$$+ 2\sum_{\ell \in [n]} \left| \underset{\vec{S}, \vec{S}' \sim \mathcal{D}^{nT}}{\mathbb{E}} \underset{Z \sim \mathcal{D}^T}{\mathbb{E}} \left[\mathbb{1} \left\{ \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| > \lambda \right\} \cdot \max_{m \in [T]} |f(S_m^{\ell-1}) - f(S_m^\ell)| \right] \right|$$

$$\leq (e^\varepsilon - e^{-\varepsilon})n\tau + 6nT\Delta.$$

$\square$

A.2. Multi Sample Amplification.

Theorem A.2 (High Probability Bound)**.** *Let $\mathcal{D}$ be a distribution over a domain X, let $f : X^n \to \mathbb{R}$, and let Δ, λ, τ be s.t. for every $1 \leq i \leq n$ it holds that*

$$\underset{\substack{S \sim \mathcal{D}^n \\ z \sim \mathcal{D}}}{\mathbb{E}} \left[\mathbb{1} \left\{ \left| f(S) - f\left(S^{(i \leftarrow z)}\right) \right| > \lambda \right\} \cdot \left| f(S) - f\left(S^{(i \leftarrow z)}\right) \right| \right] \leq \Delta,$$

and, furthermore, $\forall S \in X^n$ and $\forall 1 \leq i \leq n$ we have

$$\underset{y, z \sim \mathcal{D}}{\mathbb{E}} \left[\mathbb{1} \left\{ \left| f(S^{(i \leftarrow y)}) - f\left(S^{(i \leftarrow z)}\right) \right| \leq \lambda \right\} \cdot \left| f(S^{(i \leftarrow y)}) - f\left(S^{(i \leftarrow z)}\right) \right| \right] \leq \tau,$$

where $S^{(i \leftarrow z)}$ is the same as S except that the ith element is replaced with z. Then for every $\varepsilon > 0$ we have that

$$\Pr_{S \sim \mathcal{D}^n} \left[|f(S) - f(\mathcal{D}^n)| \geq 6(e^\varepsilon - e^{-\varepsilon})\tau n \right] < \frac{14\Delta}{(e^\varepsilon - e^{-\varepsilon})\tau},$$

provided that $n \geq O\left(\frac{\lambda}{\varepsilon(e^\varepsilon - e^{-\varepsilon})\tau} \log\left(\frac{(e^\varepsilon - e^{-\varepsilon})\tau}{\Delta} \right) \right)$

Proof. We only analyze the probability that $(f(S) - f(\mathcal{D}^n))$ is large. The analysis for $(f(\mathcal{D}^n) - f(S))$ is symmetric. Assume towards contradiction that with probability at least $\frac{7\Delta}{(e^\varepsilon - e^{-\varepsilon})\tau}$ we have that $f(S) - f(\mathcal{D}^n) \geq 6(e^\varepsilon - e^{-\varepsilon})\tau n$. We now construct the following algorithm $\mathcal{B}$ that contradicts our expectation bound.

Algorithm 2: $\mathcal{B}$

Input: T databases of size n each: $\vec{S} = (S_1, \ldots, S_T)$, where $T \triangleq \left\lfloor \frac{(e^\varepsilon - e^{-\varepsilon})\tau}{7\Delta} \right\rfloor$.

(1) Set $H = \{\bot, 1, 2, \ldots, T\}$.

(2) For $i = 1, \ldots, T$, define $q(\vec{S}, i) = f(S_i) - f(\mathcal{D}^n)$. Also set $q(\vec{S}, \bot) = 0$.

(3) Sample $t^* \in H$ with probability proportional to $\exp\left(\frac{\varepsilon}{2\lambda} q(\vec{S}, t) \right)$.

Output: t^*.

The fact that algorithm $\mathcal{B}$ is $(\varepsilon, (f, \lambda))$-differentially private follows from the standard analysis of the Exponential Mechanism of McSherry and Talwar [2007]. The proof appears in Claim A.3 for completeness.

Now consider applying $\mathcal{B}$ on databases $\vec{S} = (S_1, \ldots, S_T)$ containing i.i.d. samples from $\mathcal{D}$. By our assumption on $\mathcal{D}$ and f, for every t we have that $f(S_t) - f(\mathcal{D}^n) \geq 6(e^\varepsilon - e^{-\varepsilon})\tau n$ with probability at least $\frac{7\Delta}{(e^\varepsilon - e^{-\varepsilon})\tau}$. By our choice of $T = \left\lfloor \frac{(e^\varepsilon - e^{-\varepsilon})\tau}{7\Delta} \right\rfloor$, we therefore get

$$\Pr_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max_{t \in [T]} \{f(S_t) - f(\mathcal{D}^n)\} \geq 6(e^\varepsilon - e^{-\varepsilon})\tau n \right] \geq 1 - \left(1 - \frac{7\Delta}{(e^\varepsilon - e^{-\varepsilon})\tau} \right)^T \geq \frac{1}{2}.$$

The probability is taken over the random choice of the examples in $\vec{S}$ according to $\mathcal{D}$. Thus, by Markov's inequality,

$$\mathbb{E}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max_{t \in H} \left\{ q(\vec{S}, t) \right\} \right] = \mathbb{E}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max \left\{ 0, \max_{t \in [T]} (f(S_t) - f(\mathcal{D})) \right\} \right] \geq 3(e^\varepsilon - e^{-\varepsilon})\tau n.$$

(A.15)

So, in expectation, $\max_{t \in H} \left(q(\vec{S}, t) \right)$ is large. In order to contradict the expectation bound of Theorem A.2, we need to show that this is also the case for the index t^* that is sampled in Step 3. To that end, we now use the following technical claim, stating that the expected quality of a solution sampled as in Step 3 is high. Indeed, by the properties of the exponential mechanism (Theorem 2.5) we have that

$$
\begin{aligned}
\mathbb{E}_{t^* \in_R H}[q(\vec{S}, t^*)] &= \mathbb{E}_{t^* \in_R H} \left[\mathbb{1}\{t^* \neq \bot\} \cdot (f(S_{t^*}) - f(\mathcal{D}^n)) \right] \\
&\geq \max\{0, \max_{t \in [T]} (f(S_t) - f(\mathcal{D}^n))\} - \frac{2\lambda}{\varepsilon} \ln(T + 1).
\end{aligned}
$$

Taking the expectation also over $\vec{S} \sim \mathcal{D}^{nT}$ we get that

$$
\mathbb{E}_{\substack{\vec{S} \sim \mathcal{D}^{nT} \\ t^* \leftarrow \mathcal{B}(\vec{S})}} \left[\mathbb{1}\{t^* \neq \bot\} \cdot (f(S_{t^*}) - f(\mathcal{D}^n))\} \right]
$$

$$
\geq \mathbb{E}_{\vec{S} \sim \mathcal{D}^{nT}} \left[\max \left\{ 0 , \max_{t \in [T]} (f(S_t) - f(\mathcal{D}^n)) \right\} \right] - \frac{2\lambda}{\varepsilon} \ln(T + 1)
$$

$$
\geq 3(e^\varepsilon - e^{-\varepsilon})\tau n - \frac{2\lambda}{\varepsilon} \ln(T + 1).
$$

This contradicts Theorem A.2 whenever $n > \frac{2\lambda}{\varepsilon(e^\varepsilon - e^{-\varepsilon})\tau} \ln(T+1) = \frac{2\lambda}{\varepsilon(e^\varepsilon - e^{-\varepsilon})\tau} \ln(\frac{(e^\varepsilon - e^{-\varepsilon})\tau}{7\Delta} + 1)$.   □

Claim A.3. *Algorithm $\mathcal{B}$ is $(\varepsilon, (f, \lambda))$-differentially private.*

Proof. Fix two (f, λ)-neighboring databases $\vec{S}$ and $\vec{S'}$, and let $b \in \{\bot, 1, 2, \ldots, T\}$ be a possible output. We have that

$$
\Pr[\mathcal{B}(\vec{S}) = b] = \frac{\exp(\frac{\varepsilon}{2\lambda} \cdot q(\vec{S}, b))}{\sum_{a \in H} \exp(\frac{\varepsilon}{2\lambda} \cdot q(\vec{S}, a))} \tag{A.16}
$$

Using the fact that $\vec{S}$ and $\vec{S'}$ are (f, λ)-neighboring, for every $a \in H$ we get that $q(\vec{S'}, a) - \lambda \leq q(\vec{S}, a) \leq q(\vec{S'}, a) + \lambda$. Hence,

$$
(A.16) \leq \frac{\exp(\frac{\varepsilon}{2\lambda} \cdot [q(\vec{S'}, b) + \lambda])}{\sum_{a \in H} \exp(\frac{\varepsilon}{2\lambda} \cdot [q(\vec{S'}, a) - \lambda])}
$$

$$
= \frac{e^{\varepsilon/2} \cdot \exp(\frac{\varepsilon}{2\lambda} \cdot q(\vec{S'}, b))}{e^{-\varepsilon/2} \sum_{a \in H} \exp(\frac{\varepsilon}{2\lambda} \cdot q(\vec{S'}, a))}
$$

$$
= e^\varepsilon \cdot \Pr[\mathcal{B}(\vec{S'}) = b].
$$

For any possible *set* of outputs $B \subseteq \{\bot, 1, 2, \ldots, T\}$ we now have that

$$
\Pr[\mathcal{B}(\vec{S}) \in B] = \sum_{b \in B} \Pr[\mathcal{B}(\vec{S}) = b] \leq \sum_{b \in B} e^\varepsilon \cdot \Pr[\mathcal{B}(\vec{S'}) = b] = \Pr[\mathcal{B}(\vec{S'}) \in B].
$$

□

Appendix B. A Similar Statement of Part 1 of Theorem 1.2[1]

Let $\mathcal{D}$ be a distribution over a domain X, and let $f : X^n \to \mathbb{R}$. Our goal is to derive conditions on f and $\mathcal{D}$ s.t. w.h.p. over $S \sim \mathcal{D}^n$ we have that $|f(S) - f(\mathcal{D})|$ is small. To that end, let $S, T \sim \mathcal{D}^n$ be independent, each containing n examples from $\mathcal{D}$. We denote $S = (s_1, \ldots, s_n)$ and $T = (t_1, \ldots, t_n)$. For $i \in [n]$ we define

$$
f_i(S) = \mathbb{E}_{T \sim \mathcal{D}^n} [f(s_1, \ldots, s_i, t_{i+1}, \ldots, t_n) - f(s_1, \ldots, s_{i-1}, t_i, \ldots, t_n)].
$$

---

[1]This strategy was suggested by an anonymous reviewer of an earlier version of this paper.

We get that

$$\sum_{i=1}^{n} f_i(S) = \mathbb{E}_{T\sim\mathcal{D}^n}[f(S)] - \mathbb{E}_{T\sim\mathcal{D}^n}[f(T)] = f(S) - f(\mathcal{D}).$$

So, our goal is to show a high probability bound on $\left|\sum_{i\in[n]} f_i(S)\right|$. Let us decompose:

$$\begin{aligned}
f_i(S) &= g_i(S) + h_i(S), \qquad \text{where:} \\
g_i(S) &= \mathbb{1}_{\{|f_i(S)|\le\lambda\}} \cdot f_i(S), \\
h_i(S) &= \mathbb{1}_{\{|f_i(S)|>\lambda\}} \cdot f_i(S).
\end{aligned}$$

Observe that for every $i \in [n]$ we have that $\mathbb{E}_{S\sim\mathcal{D}^n}[f_i(S)] = 0$. Hence,

$$\begin{aligned}
\sum_{i\in[n]} f_i(S) &= \sum_{i\in[n]} \left(f_i(S) - \mathbb{E}_{S\sim\mathcal{D}^n}[f_i(S)] \right) \\
&= \underbrace{\sum_{i\in[n]} \left(g_i(S) - \mathbb{E}_{S\sim\mathcal{D}^n}[g_i(S)] \right)}_{\text{Expression 1}} + \underbrace{\sum_{i\in[n]} \left(h_i(S) - \mathbb{E}_{S\sim\mathcal{D}^n}[h_i(S)] \right)}_{\text{Expression 2}}.
\end{aligned}$$

Next we bound Expression 1 using Azuma's inequality and Expression 2 using Markov's inequality. Recall Azuma's inequality:

Theorem B.1 (Azuma's Inequality). *Let $c \ge 0$ be a constant, and let $X_1, \ldots, X_m$ be random variables s.t. for every j we have that $\mathbb{E}|X_j| < \infty$, and $\mathbb{E}[X_{j+1}|X_1, \ldots, X_j] = X_j$, and $|X_{j+1} - X_j| \le c$. Then for every $\alpha > 0$,*

$$\Pr[X_n - X_1 \ge \alpha] \le \exp\left(-\frac{\alpha^2}{2mc^2}\right).$$

Bounding Expression 1. In our case, we apply Azuma's inequality to the random variables $X_0, X_1, \ldots, X_n$, where $X_j = \sum_{i=1}^{j} (g_i(S) - \mathbb{E}_{S\sim\mathcal{D}^n}[g_i(S)])$. Indeed, for every j it holds that $|X_{j+1} - X_j| \le 2\lambda$, and

$$\mathbb{E}[X_j|X_1, \ldots, X_{j-1}] = \mathbb{E}_{S\sim\mathcal{D}^n}\left[X_{j-1} + \left(g_j(S) + \mathbb{E}_{S\sim\mathcal{D}^n}[g_j(S)] \right) \middle| X_1, \ldots, X_{j-1} \right] = X_{j-1}.$$

Hence, by Azuma's Inequality, for every $\varepsilon, \Delta > 0$ satisfying $n \ge \frac{16}{\varepsilon^2\lambda} \ln\left(\frac{\varepsilon\lambda}{\Delta}\right)$, we have

$$\Pr\left[\sum_{i\in[n]} \left(g_i(S) - \mathbb{E}_{S\sim\mathcal{D}^n}[g_i(S)] \right) \ge \varepsilon\lambda n \right] \le \exp\left(-\frac{\varepsilon^2 n}{16}\right) \le \frac{\Delta}{\varepsilon\lambda}.$$

Bounding Expression 2. In order to obtain a high probability bound on Expression 2, we make the following assumption on the function f and the distribution $\mathcal{D}$. Let $\Delta > 0$ be such that for every i we have

$$
\begin{aligned}
\Delta \; \geq \; & \underset{S \sim \mathcal{D}^n}{\mathbb{E}} |h_i(S)| \\
= \; & \underset{S \sim \mathcal{D}^n}{\mathbb{E}} \left[\mathbb{1}_{\{|f_i(S)| > \lambda\}} \cdot |f_i(S)| \right] \\
= \; & \underset{S \sim \mathcal{D}^n}{\mathbb{E}} \left[\mathbb{1}_{\left\{ |\mathbb{E}_{T \sim \mathcal{D}^n}[f(s_1, \ldots, s_i, t_{i+1}, \ldots, t_n) - f(s_1, \ldots, s_{i-1}, t_i, \ldots, t_n)]| > \lambda \right\}} \cdot \right. \\
& \qquad\qquad \left. \cdot \left| \underset{T \sim \mathcal{D}^n}{\mathbb{E}} [f(s_1, \ldots, s_i, t_{i+1}, \ldots, t_n) - f(s_1, \ldots, s_{i-1}, t_i, \ldots, t_n)] \right| \right].
\end{aligned}
$$

Using this assumption we get that

$$
\begin{aligned}
\Pr \left[\sum_{i \in [n]} \left(h_i(S) - \underset{S \sim \mathcal{D}^n}{\mathbb{E}}[h_i(S)] \right) \geq \varepsilon \lambda n + n\Delta \right] \; & \leq \; \underset{S \sim \mathcal{D}^n}{\Pr} \left[\sum_{i \in [n]} h_i(S) \geq \varepsilon \lambda n \right] \\
& \leq \; \underset{S \sim \mathcal{D}^n}{\Pr} \left[\sum_{i \in [n]} |h_i(S)| \geq \varepsilon \lambda n \right] \\
& \leq \; \frac{1}{\varepsilon \lambda n} \cdot \underset{S \sim \mathcal{D}^n}{\mathbb{E}} \left[\sum_{i \in [n]} |h_i(S)| \right] \leq \frac{\Delta}{\varepsilon \lambda}.
\end{aligned}
$$

To conclude, under our assumption on f and $\mathcal{D}$, with probability at least $1 - \frac{2\Delta}{\varepsilon \lambda}$, we have that Expression 1 is at most $\varepsilon \lambda n$ and that Expression 2 is at most $\varepsilon \lambda n + n\Delta$. That is,

Theorem B.2. *Let $\mathcal{D}$ be a distribution over a domain X, let $f : X^n \to \mathbb{R}$, and let $\Delta, \lambda \geq 0$ be s.t. for every $1 \leq i \leq n$ it holds that*

$$
\underset{S \sim \mathcal{D}^n}{\mathbb{E}} \left[\mathbb{1}_{\left\{ |\mathbb{E}_{T \sim \mathcal{D}^n}[f(s_1, \ldots, s_i, t_{i+1}, \ldots, t_n) - f(s_1, \ldots, s_{i-1}, t_i, \ldots, t_n)]| > \lambda \right\}} \cdot \right.
$$
$$
\left. \cdot \left| \underset{T \sim \mathcal{D}^n}{\mathbb{E}} [f(s_1, \ldots, s_i, t_{i+1}, \ldots, t_n) - f(s_1, \ldots, s_{i-1}, t_i, \ldots, t_n)] \right| \right] \leq \Delta.
$$

Then for every $\varepsilon > 0$ we have that

$$
\underset{S \sim \mathcal{D}^n}{\Pr} [f(S) - f(\mathcal{D}^n) \geq 2\varepsilon \lambda n + n\Delta] < \frac{2\Delta}{\varepsilon \lambda},
$$

provided that $n \geq \frac{16}{\varepsilon^2 \lambda} \ln \left(\frac{\varepsilon \lambda}{\Delta} \right)$.

Observe that Theorem B.2 is "similar in spirit" to Part 1 of Theorem 1.2 (even though they are syntactically different). We do not know if a similar analogue for exists for Part 2 of Theorem 1.2 (recall that Part 2 was needed for analyzing some of our applications in Section 4).