Differentially Private Medians and Interior Points for Non-Pathological Data

Maryam Aliakbarpour
Department of Computer Science, Rice University, Houston, TX, USA

Rose Silver
Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA

Thomas Steinke
Google DeepMind, Mountain View, CA, USA

Jonathan Ullman
Khoury College of Computer Sciences, Northeastern University, Boston, MA, USA

Abstract

We construct sample-efficient differentially private estimators for the approximate-median and interior-point problems, that can be applied to arbitrary input distributions over $\mathbb{R}$ satisfying very mild statistical assumptions. Our results stand in contrast to the surprising negative result of Bun et al. (FOCS 2015), which showed that private estimators with finite sample complexity cannot produce interior points on arbitrary distributions.

2012 ACM Subject Classification Theory of computation → Theory of database privacy and security

Keywords and phrases Differential Privacy, Statistical Estimation, Approximate Medians, Interior Point Problem

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.3

Related Version *Full Version*: `https://arxiv.org/abs/2305.13440` [1]

Funding *Maryam Aliakbarpour*: Supported by NSF grants CNS-2120667, CNS-2120603, CCF-1934846, and BU's Hariri Institute for Computing. This work was predominantly done while she was affiliated with Boston University and Northeastern University.
Rose Silver: Supported by NSF CCF-1750640 and CNS-2120603.
Jonathan Ullman: Supported by NSF CCF-1750640 and CNS-2120603.

1 Introduction

A statistical estimator is an algorithm that takes data drawn from an unknown distribution as input and tries to learn something about that distribution. While the input data is only a conduit for learning about the distribution, many statistical estimators also reveal a lot of information that is specific to the input data, which raises concerns about the *privacy* of people who contributed their data. In response, we can try to design estimators that are *differentially private (DP)* [13], which ensure that no attacker can infer much more about any person in the input data than they could have inferred in a hypothetical world where that person's data had never been collected.

Differential privacy is a strong constraint that imposes significant costs even for very simple statistical estimation tasks. In this paper we focus on two such tasks: *interior point estimation* and *median estimation*. In the interior point problem, we have a distribution P over $\mathbb{R}$, and our goal is simply to output some point y with

$$\inf \operatorname{support}(P) \leq y \leq \sup \operatorname{support}(P). \tag{1}$$

There is a trivial estimator for solving the interior point problem – draw a single sample from P and output it – but this estimator is clearly not private. More generally, we can try to find an α-approximate median of the distribution, which is a point y such that

$$\tfrac{1}{2} - \alpha \leq \Pr_{x \leftarrow P}[x \leq y] \leq \tfrac{1}{2} + \alpha. \tag{2}$$

There is also a simple estimator for computing an approximate median – draw $O(1/\alpha^2)$ samples and return the median of the samples – but this estimator also fails to be private. While these problems are nearly trivial to solve without a privacy constraint, a remarkable result of Bun, Nissim, Stemmer, and Vadhan [9] showed that there is no differentially private estimator that takes any finite number of samples and outputs even just an interior point of an arbitrary distribution. Since the interior point problem is a special case of finding an approximate median, learning threshold functions, learning halfspaces, and more, this negative result has far reaching implications.

In light of this negative result, there have been two main approaches to privately solving the interior point problem and its generalizations. The first is to assume the data comes from a finite domain, such as the integers $[T] := \{1, 2, \ldots, T\}$, in which case the optimal sample complexity is now known to be $n = \tilde{\Theta}(\log^* T)$ [5, 9, 2, 8, 15, 10]. Of course, this also suggests a heuristic for handling continuous domains by simply rounding points into a discrete domain. Such an approach can be satisfactory in settings where rounding error is tolerable, but it does not, in general, guarantee that the output is a truly valid interior point (or approximate median).[1]

The second approach, which is the approach we adopt in this paper, is to assume that the distribution satisfies some additional properties that allow us to bypass the Bun et al. lower bound. Results along this line have considered a range of assumptions such as Gaussian distributions [16], distributions with uniformly large density in a (pre-specified) neighborhood around the median [12, 18, 7], and distributions with maximum density bounded above by a known parameter [14]. These results hint at a broader theme: that, intuitively, Bun et al.'s lower bound would seem to apply only to *pathological* distributions.

In this work we propose a framework for formalizing this intuition. We show that, without knowing anything else about the distribution P, there is a very weak and broadly applicable statistical assumption – which we call *bounded normalized variance* – that is sufficient to bypass the lower bound. A distribution P with mean μ satisfies C-bounded normalized variance if

$$\frac{\mathbb{E}_{X \leftarrow P}[|X - \mu|^2]}{\mathbb{E}_{X \leftarrow P}[|X - \mu|]^2} \leq C \tag{3}$$

for some constant $C \geq 1$[2]. The way to think about distributions with $O(1)$-bounded normalized variance is that these are the distributions for which *standard deviation* is a meaningful value – that is, distributions for which the standard deviation $\sigma = \sqrt{\mathbb{E}[|X - \mu|^2]}$ serves as a constant-factor proxy for the expected absolute deviation $\mathbb{E}[|X - \mu|]$. This assumption is satisfied by most real-world inputs, as well as by essentially all natural parametric families of distributions, both discrete and continuous (see Table 1 in Section A).

[1] Depending on the distribution, it can also be difficult to decide *what granularity one should round to*. Consider, for example, a distribution with 10% of its mass at -1, with 10% of its mass at 1, and with 80% of its mass uniformly spread in $[-\kappa, \kappa]$ for some very small κ. In this case, the value of the approximate median is that it gives us an estimate for κ. However, if we use a rounding granularity that is larger than κ (e.g., if the granularity is based on the standard deviation of P), then the estimated approximate median produces no useful information.

[2] Note that by Jensen's inequality C is at least one for any distribution.

Our main theorem states that, if P satisfies $O(1)$-bounded normalized variance, then it is possible to privately compute an interior point with a sample complexity depending only on the privacy parameters ϵ and δ. This gives a strong formal sense in which the Bun et al. lower bound applies only to distributions with very unusual combinatorial structure.

▶ **Theorem 1** (Informal Statement of Theorem 8). *There is an (ε, δ)-differentially private algorithm that takes $n = \mathrm{poly}(C\varepsilon^{-1} \log \delta^{-1})$ samples from an arbitrary distribution P over $\mathbb{R}$ satisfying C-bounded normalized variance and, with high probability, returns an interior point of P.*

To understand the technical role of C-bounded normalized variance in Theorem 1, consider the following intuition. Roughly speaking, if we wish to privately compute an interior point, then we want the distribution to have two competing properties. (1) We need the samples to be *concentrated together* in order to reduce the impact of any one sample, for privacy reasons. (2) We need the samples to be *spread out*, so that we can certify for some interior point that there is a non-negligible amount of mass to both its left and right. A key insight in this paper is that C-bounded normalized variance ends up being sufficient for *both* of these purposes simultaneously – this is what makes Theorem 1 possible.

Next we turn to the problem of computing approximate medians. In the worst-case setting, finding an approximate median can actually be reduced to finding an interior point, however this reduction does not preserve the property of bounded normalized variance, so we cannot use it directly to obtain a private median algorithm. Moreover, we will see that we can turn any distribution on a bounded support into a distribution satisfying bounded normalized variance without changing the median (or approximate median), so our assumption is not sufficient to circumvent the lower bound. Nonetheless, we show that a slight (and necessary) strengthening of this assumption is enough to find an approximate median. Intuitively, this assumption is *bounded normalized variance around the median*, which means that bounded normalized variance holds even if we condition on the part of P that lies between the $\frac{1}{2} - \alpha$ and $\frac{1}{2} + \alpha$ quantiles.

▶ **Theorem 2** (Informal Statement of Theorem 18). *There is an (ε, δ)-differentially private algorithm that takes $n = \mathrm{poly}(C\varepsilon^{-1}\alpha^{-1} \log \delta^{-1})$ samples from an arbitrary distribution P over $\mathbb{R}$ satisfying C-bounded normalized variance around the median, and, with high probability, returns an α-approximate median of P.*

We note that there are many other richer tasks, such as privately learning halfspaces and privately finding a point in the convex hull in $\mathbb{R}^d$, where the best private algorithms for worst-case distributions are based on reductions to interior point or closely related problems that can be reduced to interior point [4]. Theorem 2 suggests that identifying approximate mild distributional assumptions to make these problems tractable is a fruitful direction.

1.1 Technical Overview

Privately finding an interior point

At a high level, our interior point algorithm follows the approach taken by Karwa and Vadhan [16] for finding an approximate median of a Gaussian distribution, but with a much more general analysis that allows us to rely on only weak assumptions about the distribution. First, suppose that we know the first central absolute moment of the distribution, and have rescaled the distribution so that

$$\mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|] = 1 \quad \text{and} \quad \mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|^2] \leq C,$$

where $\mu = \mathbb{E}[X]$. In this case, by Chebyshev's inequality, we know that most of the probability mass for P is not too far from the mean μ. Moreover, it cannot be that the mass is almost entirely contained in a single sub-interval of size $\leq 1/2$, as this (combined with the fact that outliers in P are relatively rare) would imply that $\mathbb{E}_{X \leftarrow P}[|X - \mu|] < 1$. Thus, if we divide the real line into an infinite set of intervals

$$\ldots, [-1, -\tfrac{1}{2}), [-\tfrac{1}{2}, 0), [0, \tfrac{1}{2}), [\tfrac{1}{2}, 1), \ldots,$$

then there will be at least two distinct intervals that contain a significant amount of mass (more than $1/\text{poly}(C)$ mass). Using standard techniques for computing differentially private histograms, we can identify two of these intervals privately, and any boundary between them must be an interior point of the distribution.

The next step is to resolve the fact that we do not know the first central absolute moment $\mathbb{E}_{X \leftarrow P}[|X - \mu|]$, and we need to privately estimate this quantity up to a small multiplicative factor. To do this, we take a set of $2n$ samples $x_1, \ldots, x_{2n}$ and create a new set of n samples $y_i = |x_{2i-1} - x_{2i}|$. Note that each of these samples y_i is sampled as $|X - X'|$ where X, X' are drawn independently from P. We will use the y_is to approximate $\mathbb{E}[|X - X'|]$, which is in turn a constant-factor approximation of $\mathbb{E}[|X - \mu|]$. Specifically, we divide $[0, \infty)$ into the infinite set of intervals

$$\ldots, [\tfrac{1}{8}, \tfrac{1}{4}), [\tfrac{1}{4}, \tfrac{1}{2}), [\tfrac{1}{2}, 1), [1, 2), [2, 4), [4, 8), \ldots$$

Using the bounded normalized variance condition, we argue that the largest interval that contains a significant amount of mass is a good approximation to the first central absolute moment (up to a $\text{poly}(C)$ factor). As before, this largest interval can be identified privately via techniques for computing differentially private histograms. Chaining this algorithm with the previous algorithm gives us our complete algorithm for finding an interior point of a distribution with bounded normalized variance.

Privately finding an approximate median

We might hope that bounded normalized variance is sufficient to also find a good α-approximate median for small α, but that is actually false. To see why, first note that the Bun et al. [9] lower bound says that there is no differentially private algorithm that can find an interior point of an arbitrary distribution, even if the distribution is supported on some bounded interval, such as $[-\frac{1}{2}, \frac{1}{2})$. Given an arbitrary distribution P on this interval, we can create a new distribution P' by adding mass at -1 with probability $1/4$ and mass at $+1$ with probability $1/4$. A simple calculation shows that this new distribution will have $O(1)$-bounded normalized variance. Moreover, any $1/5$-approximate median for P' will be an interior point of P. Since we can easily simulate access to P' using access to P, any private algorithm for computing an approximate median of a distribution with bounded normalized variance can be used to privately compute an interior point of an arbitrary distribution, which is ruled out by Bun et al. [9].

Thus, our algorithm for finding an α-approximate median requires a stronger assumption on the distribution P. First, observe that an α-approximate median of P is just any interior point of the distribution P_α that consists only of the middle 2α *slice* of the distribution P. That is, P_α is the distribution P restricted to the space between the $\frac{1}{2} - \alpha$ and $\frac{1}{2} + \alpha$ quantiles of P. We now assume that the distribution P_α has bounded normalized variance.

Intuitively, our algorithm works by finding an interior point of the distribution P_α by using a subset of our samples from P to simulate samples from P_α, but we cannot exactly generate samples from P_α without knowing the quantiles of the distribution P itself, which

is exactly what we are trying to estimate. To get around this issue, we instead take a collection of n samples, then sort them to obtain $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, and then use the middle $(1 - 1/k)2\alpha n$ samples $\mathbf{x_0}$ as an approximation of samples from P_α (where k is some well-chosen quantity). The $1 - 1/k$ term is to ensure that the samples in $\mathbf{x_0}$ come from *within* P_α, instead of being in a situation where a small number of them may come from outside of it.

While this collection $\mathbf{x_0}$ of samples does not have the same distribution as i.i.d. samples from P_α, we are nonetheless able to argue that they come from a distribution with $O(C)$-bounded normalized variance – this is the main technical component of our approximate-median result (see Lemma 20 and Claim 21). It follows that our interior point algorithm will succeed in identifying an interior point of P_α, and thus in identifying an α-approximate median of P.

1.2 Related Work

As discussed earlier, there are two lines of prior work on privately computing interior points. The first – and most common – approach is to assume a discrete domain of size T and make no other assumptions. After a long line of work on both lower and upper bounds [5, 9, 2, 8, 15, 10], it is now known that the optimal private sample complexity, as a function of T, is $n = \tilde{\Theta}(\log^* T)$.

The most closely related prior work to ours makes distributional assumptions and otherwise allows the data to be arbitrary real numbers. Karwa and Vadhan [16] give an estimator which is specific to Gaussian distributions. Our work can be viewed as an extension of their algorithmic approach, but with an analysis that extends to any distribution satisfying the bounded normalized variance assumption. A different approach, taken by Dwork and Lei [12], Tzamos, Vlatakis-Gkaragkounis, and Zadik [18], Avella-Medina and Brunel [7], and Asi and Duchi [3], is to rely on some assumptions about the density of the probability distribution around the median. Specifically, they assume that the probability density is lower bounded at every point in some fixed-size interval around the median. Under this assumption the empirical median is a very well-behaved estimate of the true median and, as a result, one can privately estimate it using techniques based on local sensitivity. Finally, a third approach, taken by Haghtalab, Roughgarden, and Shetty [14] is to assume that the distribution has a certain *smoothness property* everywhere – namely, that the probability density has some known upper bound $1/\gamma$ that holds at all points. Intuitively, this closes the gap between continuous and discrete distributions by ensuring that there is a natural (and known) discretization granularity (bins of size $O(\gamma)$) at which the distribution is guaranteed to behave well.

The three directions of work discussed above combine to tell a compelling story: that, intuitively, the distributions that Bun et al.'s $\Omega(\log^* T)$ lower bound applies to are those with very unusual pathological structures. At the same time, there are natural distributions that are not captured in any of the three models.[3] The main goal of our work is to offer a single framework that captures almost all natural distributions, and that establishes a general-purpose mechanism for bypassing the $\Omega(\log^* T)$ lower bound.

[3] Consider, for example, the distribution P obtained by: selecting a uniformly random $x \in [0, \gamma^2/2]$ (here, $\gamma \in (0, 1)$ is the same parameter as used earlier), and returning one of either either $x + 1$ or $-(x + 1)$ at random. It is straightforward to verify that this distribution satisfies none of the three properties discussed above, although, of course, it does have $O(1)$-bounded normalized variance.

2 Preliminaries

Let P be a data distribution. We indicate that a data point x is drawn from P by writing $x \leftarrow P$. We indicate that $\mathbf{x} = (x_1, \ldots, x_n)$ is a set of n i.i.d. data points drawn from P by writing $\mathbf{x} \leftarrow P^n$. We refer to $\mathbf{x}$ as a *dataset*. We use $\mathbb{I}_\varphi$ to denote the indicator random variable for the property φ.

2.1 Differential Privacy

We say that two datasets $\mathbf{x}$ and $\mathbf{x}'$ are *neighboring datasets* if they differ in at most one data point, i.e. $\mathrm{D_H}(\mathbf{x}, \mathbf{x}') \leq 1$ where $\mathrm{D_H}$ denotes the Hamming distance.

▶ **Definition 3.** *Let $\varepsilon > 0$, $0 < \delta < 1$. An algorithm $\mathcal{A} : \mathcal{X}^n \to \mathcal{Y}$ is (ε, δ)-differentially private if, for every $E \subseteq \mathcal{Y}$ and neighboring datasets $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^n$, $\mathcal{A}$ satisfies* $\Pr[\mathcal{A}(\mathbf{x}) \in E] \leq \mathrm{e}^\varepsilon \Pr[\mathcal{A}(\mathbf{x}') \in E] + \delta$.

▶ **Definition 4.** *Let $f : \mathcal{X}^n \to \mathbb{R}^d$. The global sensitivity Δ of f is defined as* $\Delta \coloneqq \sup_{\substack{\mathbf{x},\mathbf{x}' \in \mathcal{X}^n \\ \mathrm{D_H}(\mathbf{x},\mathbf{x}')=1}} ||f(\mathbf{x}) - f(\mathbf{x}')||_1$.

The truncated Laplace mechanism

We make use of the standard approach of adding noise proportional to the global sensitivity [13] to ensure differential privacy. Since we are interested in adding noise to a histogram with infinitely many bins, we need to make use of the *truncated Laplace distribution* rather than the standard Laplace distribution. Given parameters $\lambda, Z_{\max}$, we define the truncated Laplace distribution $\mathrm{TLap}(\lambda, Z_{\max})$ over the support $[-Z_{\max}, Z_{\max}]$ with density $f(z) \propto e^{-|z|/\lambda}$.

▶ **Lemma 5** (Truncated Laplace Mechanism). *Let $f : \mathcal{X}^n \to \mathbb{R}$ be a function with global sensitivity Δ. For every $\varepsilon, \delta \in (0, 1)$, if Z_{max} is at least $\Delta \ln(4/\delta)/\varepsilon$, then the truncated Laplace mechanism $M(\mathbf{x}) \coloneqq f(\mathbf{x}) + Z$, where $Z \leftarrow \mathrm{TLap}(\frac{\Delta}{\varepsilon}, Z_{max})$ is (ε, δ)-differentially private.*

While the above lemma is considered folklore in the field, we include a proof in the full version of the paper. A standard application of this mechanism is computing histograms, possibly in infinite dimension (see e.g. [19]).

▶ **Lemma 6** (Differentially Private Histograms). *Let $\mathcal{X}$ be a domain and let $\mathcal{X}_1, \ldots, \mathcal{X}_m$ be a partition of the domain into (a possibly infinite number of) bins. Define the function $f : \mathcal{X}^n \to \mathbb{R}^m$ as $f(\mathbf{x})_j = \sum_{i=1}^n \mathbb{I}_{x_i \in \mathcal{X}_j}$. Then the mechanism*

$$M(\mathbf{x}) \coloneqq f(\mathbf{x}) + (Z_1, \ldots, Z_m),$$

where each $Z_j \leftarrow \mathrm{TLap}\left(\frac{4}{\varepsilon}, \frac{8\ln(8/\delta)}{\varepsilon}\right)$ is (ε, δ)-differentially private.

2.2 Problem Definitions

The interior point problem

Let P be a distribution over $\mathbb{R}$, and let $\mathbf{x}$ be an n-dimensional dataset. As noted earlier, we define an *interior point of P* to be any point y satisfying $\inf \mathrm{support}(P) \leq y \leq \sup \mathrm{support}(P)$. Similarly, we define an *interior point of* $\mathbf{x}$ to be any point y satisfying $\min_{x_i \in \mathbf{x}} x_i \leq y \leq \max_{x_i \in \mathbf{x}} x_i$.

The interior point problem is to compute an interior point of P when given access to $\mathbf{x} \leftarrow P^n$ for an unknown P. Observe that, when $\mathbf{x} \leftarrow P^n$, any interior point for $\mathbf{x}$ is guaranteed to also be an interior point of P.

Approximate medians

For a distribution P, let $F_P(x) \coloneqq \Pr_{X \leftarrow P}[X \leq x]$ denote the CDF of P. We use $Q_P(p) \coloneqq F_P^{-1}(p)$ to denote the p-th quantile of P for any $p \in [0,1]$. That is, $Q_P(p) = \inf\{x \mid F_P(x) \geq p\}$ for $p \in [0,1]$. For a dataset $\mathbf{x}$, we can similarly define $Q_{\mathbf{x}}(p) \coloneqq x_{(\lfloor pn \rfloor)}$ for $p \in [1/n, 1]$. We say that $\hat{m}$ is an α-approximation to the median $Q_P(0.5)$ if $|F_P(\hat{m}) - 0.5| \leq \alpha$.

For the interior point problem, we focus on data distributions with the following property:

▶ **Definition 7.** *Let P be a distribution with mean $\mu = \mathbb{E}_{X \leftarrow P}[X]$. The distribution P has C-*bounded normalized variance *for some value C if*

$$\frac{\mathbb{E}_{X \leftarrow P}[|X - \mu|^2]}{\mathbb{E}_{X \leftarrow P}[|X - \mu|]^2} \leq C.$$

*As shorthand, we will sometimes say simply that such a distribution is C-*bounded*.*

For the approximate median problem, we will be interested in the case where the *middle* 2α*-percentile* of the distribution is C-bounded, rather than the entire distribution itself.

3 An Algorithm for Interior Points

In this section, we introduce an algorithm (Algorithm 1) that privately solves the interior point problem when the data points are coming from a C-bounded distribution. Recall that, given a dataset $\mathbf{x}$, the goal of such an algorithm is to privately output a point y that falls within the minimum and maximum values in $\mathbf{x}$. Formally, we show that Algorithm 1 is an (ε, δ)-differentially private algorithm that, if P is C-bounded, returns an interior point of $\mathbf{x}$ with probability at least $1 - \beta$ if the size of $\mathbf{x}$ is sufficiently large (depending on the values of β, ε, δ, and C).

The basic idea is to apply bounded normalized variance to a private histogram. In particular, the domain of P is partitioned into contiguous bins B of a fixed width, and each bin counts the number of samples from $\mathbf{x}$ that reside in the corresponding subset of the domain. Then, random truncated Laplace noise is added to the count of every bin to ensure privacy. In essence, every bin keeps a noisy count of the number of points $x \in \mathbf{x}$ which land in the bin.

To demonstrate the benefit of this private histogram, suppose that there are two bins B_1 and B_2 in the domain of P, each of which has a sufficiently high noisy count. In particular, if the counts are high enough, then each of bins B_1 and B_2 must contain at least one $x \in \mathbf{x}$ (i.e., the large counts cannot be created entirely by the truncated Laplace noise). Moreover, any point in the domain between B_1 and B_2 *must* be an interior point of $\mathbf{x}$. The convenience of this observation is that, even though the exact locations of $\mathbf{x}$ in each of B_1 and B_2 are unknown (and even the exact number of points in B_1 and B_2 is only known up to truncated Laplace noise), it is still possible to return an interior point.

There are two possible failure modes that the algorithm might incur. The first is that the points in $\mathbf{x}$ are so spread out that no bin contains very many samples. The second is that the samples in $\mathbf{x}$ are so tightly concentrated, that *only one* bin contains a large number of samples. In order for the algorithm to succeed, we need to ensure that *at least two bins* contain a significant number of samples.

The key insight is that, if P is C-bounded, and if the bin width is chosen in the right way, then the algorithm is guaranteed to succeed with probability at least $1 - \beta$. It turns out that, in order to choose the appropriate bin width, one must first compute an estimate for the first central absolute moment of P – this is performed by a subroutine `estimate-first-moment` which, as we shall discuss in Section 3.1, *also* exploits the C-boundedness of P. By using a bin width that is slightly smaller than the first-moment estimate produced by `estimate-first-moment`, we are able to argue that at least two bins will have high noisy counts, and hence that the algorithm will succeed.

We present the guarantees of Algorithm 1 in the following theorem, whose proof is given at the end of the section.

▶ **Theorem 8.** *Suppose we are given four parameters $\epsilon > 0$, $\delta \in (0, 1)$, $\beta \in (0, 1)$, and $C > 2$. Algorithm 1 is (ϵ, δ)-differentially private. Furthermore, if P has C-bounded normalized variance, and $\mathbf{x} \leftarrow P^n$ contains n data points where*

$$n > k_0 C^3 \sqrt{\log C} \cdot \left(\epsilon^{-1} \ln \delta^{-1} + \ln \beta^{-1}\right) \tag{4}$$

for some sufficiently large positive constant k_0, then Algorithm 1 returns an interior point of $\mathbf{x}$ with probability at least $1 - \beta$.

We remark that, in the statement of Algorithm 1, as well as in the analyses of the algorithm, we have opted to give explicit constants for concreteness. Note that we have not made any effort to optimize the constants involved, but with more careful bookkeeping, these constants could almost certainly be made much smaller.

3.1 Privately Estimating the First Central Absolute Moment

In this section, we introduce `estimate-first-moment`, a private algorithm for estimating the first central absolute moment of the data distribution P, up to a multiplicative factor of $O(C\sqrt{\log C})$. The first central absolute moment approximately measures how much a random variable deviates from its mean on average. More formally, the first central absolute moment is defined as $\mathbb{E}_{X \leftarrow P}[|X - \mu|]$ where $\mu \coloneqq \mathbb{E}_{X \leftarrow P}[X]$.

Privately estimating first central absolute moments *without* estimating μ

In order to calculate the first central absolute moment of P, it would be helpful to have a good approximation of μ. Unfortunately, it is hard to *privately* calculate a good approximation to μ when the samples are unbounded; any function that averages samples together would have unbounded sensitivity, meaning that an enormous amount of noise would need to be added in order to maintain privacy.

Instead, we consider another strategy for estimating the first central absolute moment of P. For independent $X, X' \leftarrow P$ let Q be a random variable that indicates the difference of X and X': $Q \coloneqq |X - X'|$. The random variable Q is advantageous for directly estimating the first central absolute moment of P. This is in part due to the expected value $\mathbb{E}[Q]$ being a good proxy to the first central absolute moment of P, as shown by Lemma 10. Moreover, we will see that the distribution of Q enables us to privately calculate $\mathbb{E}[Q]$.

Overview of the algorithm

The algorithm `estimate-first-moment` estimates $E[Q]$ and uses it as a proxy for the first central absolute moment of P. It takes as input $\mathbf{x} \leftarrow P^n$ and extracts samples $\mathbf{q} \leftarrow Q^{n/2}$. It then creates a histogram over the domain of Q, consisting of contiguous bins whose widths are increasing powers of 2. Each bin maintains a count of the number of samples $q \in \mathbf{q}$ that

Algorithm 1 Interior Point Algorithm.

Function `estimate-first-moment`$(\mathbf{x}; \varepsilon, \delta, C)$**:**

Set $k' = 3000$ and $n = |\mathbf{x}|$.
Set $\mathbf{q} = (q_1, \ldots, q_{n/2})$, where $q_i = |x_{2i} - x_{2i-1}|$.
For all $\ell \in \mathbb{Z}$, set $c_\ell(\mathbf{q}) = \#\{q_i \mid q_i \in (2^\ell, 2^{\ell+1}]\}$ and $\hat{c}_\ell(\mathbf{q}) = c_\ell(\mathbf{q}) + Z_\ell$, where each $Z_\ell \leftarrow \mathrm{TLap}(\frac{8}{\varepsilon}, \frac{16\ln(16/\delta)}{\varepsilon})$ independently.
Set $S = \{\ell \mid \hat{c}_\ell(\mathbf{q}) \geq 3n/(8k'C\log C)\}$.
if $|S| \geq 1$ **then**
  return $\max_{\ell \in S} 2^{\ell+1}$.
else
  return $\perp$.

Function `find-interior-point`$(\mathbf{x}; \varepsilon, \delta, C, \hat{m})$**:**

Set $k' = 3000$, $k = 4096k'$, and $n = |\mathbf{x}|$.
For all $\ell \in \mathbb{Z}$, set $B_\ell = [\ell\hat{m}/\left(2k'C\sqrt{\log C}\right), (\ell+1)\hat{m}/\left(2k'C\sqrt{\log C}\right))$.
For all $\ell \in \mathbb{Z}$, set $c_\ell(\mathbf{x}) = \#\{x_i \mid x_i \in B_\ell\}$ and $\hat{c}_\ell(\mathbf{x}) = c_\ell(\mathbf{x}) + Z_\ell$, where each $Z_\ell \leftarrow \mathrm{TLap}(\frac{8}{\varepsilon}, \frac{16\ln(16/\delta)}{\varepsilon})$ independently.
Set $S = \left\{\ell \mid \hat{c}_\ell(\mathbf{x}) \geq \frac{3n}{kC^3\sqrt{\log C}}\right\}$.
if $|S| \geq 2$ **then**
  return $\frac{1}{2}\left(\min_{\ell \in S} \frac{\ell\hat{m}}{\left(2k'C\sqrt{\log C}\right)} + \max_{\ell \in S} \frac{(\ell+1)\hat{m}}{\left(2k'C\sqrt{\log C}\right)}\right)$.
else
  return $\perp$.

Function `interior-point-main`$(\mathbf{x}; \varepsilon, \delta, C)$**:**

$\hat{m} \leftarrow$ `estimate-first-moment`$(\mathbf{x}; \varepsilon, \delta, C)$.
if $\hat{m} \neq \perp$ **then**
  return `find-interior-point`$(\mathbf{x}; \varepsilon, \delta, C, \hat{m})$.
else
  return $\perp$.

land in the bin, and truncated Laplace noise is added to each count, to maintain privacy. The algorithm then eliminates all bins with small counts. Finally, the algorithm finds the largest of the remaining bins and outputs a fence post of this bin. Critically, the correctness of this algorithm will again rely heavily on the fact that P is C-bounded.

To understand why `estimate-first-moment` returns a good estimate of the first absolute moment of P, it helps to focus on the distribution of values for Q. We show that the larger values of Q appear with low probability. In particular, values of Q more than tC times larger than $\mathbb{E}[Q]$ have probability that drops as a function of $1/t^2$, as evidenced by Lemma 11. Thus, the really wide bins (which are simultaneously the bins very far away from $\mathbb{E}[Q]$) will not have much probability mass in expectation and will be eliminated. At the same time, the bins that live very close to $E[Q]$ will receive a large fraction of the mass in expectation, as evidenced by Lemma 12. Since `estimate-first-moment` returns the fence post of the widest bin of those remaining after elimination, the algorithm is likely to return a point in the domain of Q close to $\mathbb{E}[Q]$. We give the exact details of the performance of `estimate-first-moment` in Proposition 9.

Notation

Assume the data distribution P is C-bounded, and let $\mu = \mathbb{E}_{X \leftarrow P}[X]$ denote the mean of P. Our privacy parameters are ϵ and δ. Our confidence parameter is β: that is, the algorithm outputs the appropriate answer with probability at least $1 - \beta$.

▶ **Proposition 9.** *Let $\varepsilon > 0$, $\delta \in (0, 1)$, and $\beta \in (0, 1)$. Let P be a C-bounded distribution for $C > 2$. Let $\mathbf{x} \leftarrow P^{2n}$ be a dataset of $2n$ data points from P where n satisfies*

$$n \geq k\, C \log(C) \left(\ln(2/\beta) + \frac{16 \ln(16/\delta)}{\varepsilon} \right),$$

for a sufficiently large constant k. Then, there exists a constant k' such that `estimate-first-moment(x`$;\varepsilon, \delta, C$`)` *returns an estimate $\hat{m}$ for which the following guarantee holds with probability at least $1 - \beta$:*

$$\mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|] \leq \hat{m} \leq \left(2k' C \sqrt{\log C} \right) \mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|]\,. \tag{5}$$

To prove Proposition 9, we first need to introduce three useful lemmas. The first lemma (Lemma 10) states that $\mathbb{E}[Q]$ is within a multiplicative factor of 2 of the first central absolute moment. The proof of this lemma is deferred to the full version of the paper.

▶ **Lemma 10.** *Let X and X' be two random variables independently drawn from P with mean μ, and let Q be a random variable that indicates the difference of X and X': $Q \coloneqq |X - X'|$. Then, we have*

$$\mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|] \leq \mathbb{E}[Q] \leq 2 \mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|]\,.$$

As previously mentioned, the bins in the private histogram that are very far away from $\mathbb{E}[Q]$ are highly likely to be eliminated by `estimate-first-moment`. We demonstrate this with the second lemma (Lemma 11), whose proof is deferred to the full version of the paper.

▶ **Lemma 11.** *Let X and X' be two random variables independently drawn from a distribution P, and let Q be a random variable that indicates the absolute difference of X and X'. Suppose P is C-bounded for some $C \geq 1$. For any $t > 0$, we have*

$$\Pr\left[Q - \mathbb{E}[Q] \geq tC\, \mathbb{E}[Q]\right] \leq \frac{4}{t^2 C}.$$

For the third lemma (Lemma 12), consider the interval $\mathcal{I} = \left[\frac{1}{2}\mathbb{E}[Q], k'C\sqrt{\log C}\,\mathbb{E}[Q]\right]$ which surrounds $\mathbb{E}[Q]$ (k' is a constant). We show that there exists a bin in this range that receives a high count in expectation. The proof of this lemma is deferred to the full version of the paper.

▶ **Lemma 12.** *Let $k' = 3000$, $C > 2$. If P is C-bounded, then there is some ℓ satisfying*

$$(2^\ell, 2^{\ell+1}] \subseteq \mathcal{I} \text{ and } \Pr\left[Q \in (2^\ell, 2^{\ell+1}]\right] \geq \frac{1}{k' C \log C}.$$

We now give the proof of Proposition 9.

Proof (Proposition 9). Recall that `estimate-first-moment` first creates points $q_1, \ldots, q_n$ and then uses these points to realize a noisy histogram over intervals of the form $(2^\ell, 2^{\ell+1}]$. It then identifies all intervals with $\hat{c}_\ell(\mathbf{q})$ larger than the threshold $3n/(8k'C \log C)$. Of these intervals, it chooses the largest ℓ and outputs $\hat{m} = 2^{\ell+1}$ for this ℓ. If this largest interval $(2^\ell, 2^{\ell+1}]$ satisfies $(2^\ell, 2^{\ell+1}] \subseteq \mathcal{I}$, then by Lemma 10, $\hat{m} = 2^{\ell+1}$ satisfies (5).

We now turn our attention towards the two ways in which `estimate-first-moment` can fail to output an estimate $\hat{m}$ satisfying (5). The first mode of failure occurs if there is no such ℓ such that $(2^\ell, 2^{\ell+1}] \subseteq \mathcal{I}$ and $\hat{c}_\ell(\mathbf{q}) \geq 3n/(8k'C\log C)$. In particular, we can define E_1 to be the event that for all ℓ such that $(2^\ell, 2^{\ell+1}] \subseteq I$, $\hat{c}_\ell(\mathbf{q}) < 3n/(8k'C\log C)$. The second mode of failure occurs if the output $2^{\ell+1}$ is too large; in particular, we can define E_2 to be the event that there exists an ℓ such that $(2^\ell, 2^{\ell+1}] \subseteq [k'C\sqrt{\log C}\,\mathbb{E}[Q], \infty)$ and $\hat{c}_\ell(\mathbf{q}) > 3n/(8k'C\log C)$. The following two lemmas bound the probability of these bad events occurring.

▶ **Lemma 13.** *Let $\varepsilon > 0$, $\delta \in (0,1)$, $\beta \in (0,1)$, and let $C > 2$ be parameters. Let $k' = 3000$ and $k = 8k'$. Let $\mathbf{x} \leftarrow P^{2n}$ be the samples fed into the algorithm* `estimate-first-moment`*. If we have both that P is C-bounded and that $n \geq kC\log(C)\left(\ln(2/\beta) + 16\ln(16/\delta)/\varepsilon\right)$, then $\Pr[E_1] \leq \beta/2$, where the probability is taken over both the randomness of the samples $\mathbf{x}$ and the truncated Laplace mechanism.*

Proof. Let ℓ^* be the $\arg\max \hat{c}_\ell(\mathbf{q})$ over all choices of ℓ such that $(2^\ell, 2^{\ell+1}] \subseteq \mathcal{I}$. Lemma 12 implies that

$$\mathop{\mathbb{E}}_{\mathbf{q}}[c_{\ell^*}(\mathbf{q})] = n\Pr\left[Q \in (2^{\ell^*}, 2^{\ell^*+1}]\right] \geq \frac{n}{k'C\log C}. \tag{6}$$

Thus, one can bound the probability of E_1 as follows:

$$\begin{aligned}
\Pr_{\mathbf{q},Z_\ell}[E_1] &= \Pr\left[\hat{c}_{\ell^*}(\mathbf{q}) < \frac{3n}{8k'C\log C}\right] = \Pr\left[c_{\ell^*}(\mathbf{q}) + Z_{\ell^*} < \frac{3n}{8k'C\log C}\right] \\
&\leq \Pr\left[c_{\ell^*}(\mathbf{q}) < \frac{3n}{8k'C\log C} + \frac{16\ln(16/\delta)}{\varepsilon}\right] \\
&\leq \Pr\left[c_{\ell^*}(\mathbf{q}) < \frac{3n}{8k'C\log C} + \frac{n}{8k'C\log C}\right] && \text{(by assumption on } n\text{)} \\
&\leq \Pr\left[c_{\ell^*}(\mathbf{q}) < \frac{\mathbb{E}[c_{\ell^*}(\mathbf{q})]}{2}\right]. && \text{(by (6))}
\end{aligned}$$

Let Q_j be the indicator random variable for whether $q_j \in (2^{\ell^*}, 2^{\ell^*+1}]$ and note that $c_{\ell^*}(\mathbf{q}) = \sum_{j=1}^n Q_j$. Thus, by a Chernoff bound, we get:

$$\begin{aligned}
\Pr\left[c_{\ell^*}(\mathbf{q}) < \frac{\mathbb{E}[c_{\ell^*}(\mathbf{q})]}{2}\right] &\leq \exp\left(-\frac{\mathbb{E}[c_{\ell^*}(\mathbf{q})]}{8}\right) \leq \exp\left(-\frac{n}{8k'C\log C}\right) && \text{(by (6))} \\
&\leq \exp\left(-\ln(2/\beta)\right) = \beta/2\,. && \text{(by assumption on } n\text{)}
\end{aligned}$$

◀

▶ **Lemma 14.** *Let $\varepsilon > 0$, $\delta \in (0,1)$, and $\beta \in (0,1)$ be parameters. If P is C-bounded for some parameter $C > 2$, and $n \geq kC\log C\left(\ln(2/\beta) + 16\varepsilon^{-1}\ln(16/\delta)\right)$, then $\Pr[E_2] \leq \beta/2$, where the probability is taken over both the randomness of the samples $\mathbf{x}$ and the truncated Laplace mechanism.*

We prove this lemma via an approach analogous to Lemma 13. For the proof of this lemma, see the full version of the paper.

The algorithm `estimate-first-moment` fails to output the desired estimate $\hat{m}$ if either E_1 and/or E_2 occur. Lemma 13 tells us that $\Pr[E_1] \leq \beta/2$, and Lemma 14 tells us that $\Pr[E_2] \leq \beta/2$. Thus, by a union bound,

$$\Pr[E_1 \cup E_2] \leq \Pr[E_1] + \Pr[E_2] \leq \beta$$

which implies the proposition. ◀

3.2 Finding an Interior Point, Given a Fixed Bin Width

In this section, we give a guarantee on the success probability of `find-interior-point`. Recall that the algorithm instantiates a histogram, counting the number of samples in $\mathbf{x}$ that fall into sets of contiguous bins (where the width of each bin is slightly smaller than the output of `estimate-first-moment`). The algorithm adds truncated Laplace noise to the count of each bin, ensuring that the histogram is private. Then, the algorithm isolates all bins with large counts. Of all the isolated bins, the algorithm picks two and finally returns a value which falls between the domains of each of the two bins.

Bounded normalized variance induces multiple full bins

If the algorithm is able to identify multiple bins that each have samples from $\mathbf{x}$, then the algorithm is guaranteed to succeed. The C-boundedness assumption on the data distribution guarantees the existence of at least two such bins with high probability (at least $1-\beta$) over the samples.

The basic idea behind the analysis is as follows. If there is a large probability mass concentrated in a single bin (but not in any others), then we would be able to use C-boundedness in order to deduce that the true first central absolute moment of P is actually much smaller than our bin size – this would contradict Lemma 14. On the other hand, if P's probability mass is so spread out that *no bin* is expected to contain a large noisy count, then we could use C-boundedness in order to argue that P violates Chebyshev's inequality, again leading to a contradiction. Thus we are able to conclude (in Lemma 17) that at least two bins should have large noisy counts (with high probability). The full guarantees provided by `find-interior-point` are laid out in Proposition 15.

▶ **Proposition 15.** *Let $\varepsilon > 0$, $\delta \in (0,1)$, $\beta \in (0,1)$, $k' = 3000$, $k = 4096k'$. Let P be a C-bounded distribution for $C > 2$ and mean μ, and let $\mathbf{x} \sim P^n$ for*

$$n \geq kC^3\sqrt{\log C}\left(\log(2/\beta) + \frac{16\ln(16/\delta)}{\varepsilon}\right).$$

If $\mathbb{E}_{X\leftarrow P}[|X-\mu|] \leq \hat{m} \leq \left(2k'C\sqrt{\log C}\right)\mathbb{E}_{X\leftarrow P}[|X-\mu|]$, then the algorithm `find-interior-point`$(x;\varepsilon,\delta,C,\hat{m})$ *returns an interior point of $\mathbf{x}$ with probability at least $1-\beta$.*

Proof (Proposition 15). To simplify notation, we let $Z = |X-\mu|$ for $X \leftarrow P$. We start by turning our attention to the set S, introduced in `find-interior-point`, which stores the indices to bins with counts above a desirable threshold. Observe that S branches `find-interior-point` into two cases: either $|S| \geq 2$ or $|S| < 2$. To analyze these two cases, we must begin by making the following claim about S and consequently $c_\ell(\mathbf{x})$, the non-noisy count of each bin:

▷ Claim 16. For all $\ell \in S$, we have that $c_\ell(\mathbf{x}) > 0$.

Proof (Claim 16). By construction, $\ell \in S$ if $\hat{c}_\ell(\mathbf{x}) > \frac{3n}{kC^3\sqrt{\log C}}$. Since $\hat{c}_\ell(\mathbf{x}) = c_\ell(\mathbf{x}) + Z_\ell$, this implies that

$$c_\ell(\mathbf{x}) > \frac{3n}{kC^3\sqrt{\log C}} - Z_\ell \geq \frac{3n}{kC^3\sqrt{\log C}} - \frac{16\ln(16/\delta)}{\varepsilon} > 0$$

by the assumption on n. ◁

In the case where $|S| \geq 2$, the algorithm will always return an interior point. This is because the algorithm picks two $\ell_1, \ell_2 \in S$ and outputs a point p in the domain that lies between B_{ℓ_1} and B_{ℓ_2}. By Claim 16, we know that B_{ℓ_1} and B_{ℓ_2} each receive at least one sample each from $\mathbf{x}$, and so p must be an interior point of $\mathbf{x}$.

In the case where $|S| < 2$, the algorithm will always fail to output an interior point (since the algorithm defaults to $\perp$ in this case). Thus, we prove the proposition by showing that $|S| < 2$ with probability at most β.

To analyze the probability that $|S| < 2$, we need to look at the distribution P. It turns out that, if P is C-bounded for some known $C > 1$, we are guaranteed that there exists two disjoint regions, at most a distance $\mathbb{E}[Z]/2$ apart, that each contain support in P. In particular, we have the following lemma, whose proof is deferred to the full version of the paper:

▶ **Lemma 17.** *Suppose P is C-bounded for some known $C \geq 1$. Let $k_1 \geq 2$. Then*

$$\Pr\left[X \in \left(\mu + \frac{\mathbb{E}[Z]}{2k_1}, \mu + 16C\,\mathbb{E}[Z]\right)\right] \geq \frac{1}{128C} \tag{7}$$

and

$$\Pr\left[X \in \left(\mu - 16C\,\mathbb{E}[Z], \mu - \frac{\mathbb{E}[Z]}{2k_1}\right)\right] \geq \frac{1}{128C}. \tag{8}$$

This implies that, there exists two disjoint intervals B_{ℓ_1} and B_{ℓ_2} with support in P. If $|S| < 2$, then at least one of these two intervals did not receive any samples from $\mathbf{x}$, and either $\ell_1 \notin S$ or $\ell_2 \notin S$. We begin by lower bounding the expected number of samples in B_{ℓ_1}, i.e. $\mathbb{E}[c_{\ell_1}(\mathbf{x})]$. Lemma 17 tells us that

$$\Pr_{X \leftarrow P}\left[X \in \left(\mu + \frac{\mathbb{E}[Z]}{2k_1}, \mu + 16C\,\mathbb{E}[Z]\right)\right] \geq \frac{1}{128C}.$$

The size of each interval B_ℓ is $\hat{m}/\left(2k'C\sqrt{\log C}\right)$, and the size of the interval $\left(\mu + \frac{\mathbb{E}[Z]}{2k_1}, \mu + 16C\,\mathbb{E}[Z]\right)$ is at most $16C\,\mathbb{E}[Z]$. Thus, the number of intervals B_ℓ within $\left(\mu + \frac{\mathbb{E}[Z]}{2k_1}, \mu + 16C\,\mathbb{E}[Z]\right)$ is at most

$$\frac{16C\,\mathbb{E}[Z]}{\hat{m}/\left(2k'C\sqrt{\log C}\right)} \leq \frac{16C\,\mathbb{E}[Z]\left(2k'C\sqrt{\log C}\right)}{\mathbb{E}[Z]} \leq 16C\left(2k'C\sqrt{\log C}\right).$$

This implies that there exists an ℓ_1 such that $B_{\ell_1} \subseteq \left(\mu + \frac{\mathbb{E}[Z]}{2k_1}, \mu + 16C\,\mathbb{E}[Z]\right)$ and

$$\mathbb{E}[c_{\ell_1}(\mathbf{x})] \geq \frac{n}{128C} \cdot \frac{1}{16C\left(2k'C\sqrt{\log C}\right)} = \frac{n}{kC^3\sqrt{\log C}}.$$

Thus, it follows that

$$\begin{aligned}
\Pr_{\mathbf{x} \leftarrow P^n, Z_\ell}\left[\hat{c}_{\ell_1}(\mathbf{x}) < \frac{3n}{kC^3\sqrt{\log C}}\right] &= \Pr\left[c_{\ell_1}(\mathbf{x}) + Z_\ell < \frac{3n}{kC^3\sqrt{\log C}}\right] \\
&\leq \Pr\left[c_{\ell_1}(\mathbf{x}) < \frac{3n}{kC^3\sqrt{\log C}} + \frac{16\ln(16/\delta)}{\varepsilon}\right] \\
&\leq \Pr\left[c_{\ell_1}(\mathbf{x}) < \frac{3n}{kC^3\sqrt{\log C}} + \frac{n}{kC^3\sqrt{\log C}}\right] \\
&\qquad \text{(by the assumption on } n\text{)}
\end{aligned}$$

$$
\begin{aligned}
&\leq \Pr\left[c_{\ell_1}(\mathbf{x}) < \frac{\mathbb{E}[c_{\ell_1}(\mathbf{x})]}{2}\right] \\
&\leq \exp\left(-\frac{\mathbb{E}[c_{\ell_1}(\mathbf{x})]}{8}\right) && \text{(by a Chernoff bound)} \\
&\leq \exp\left(-\frac{\ln(2/\beta)}{8}\right) && \text{(by the assumption on } n\text{)} \\
&\leq \frac{\beta}{2}.
\end{aligned}
$$

By symmetry, we can also show that there exists an ℓ_2 satisfying both $B_{\ell_2} \subseteq \left(\mu - 16C\,\mathbb{E}[Z], \mu - \frac{\mathbb{E}[Z]}{2k_1}\right)$ and $\Pr\left[\hat{c}_{\ell_2}(\mathbf{x}) < \frac{3n}{kC^3\sqrt{\log C}}\right] \leq \beta/2$. Putting the pieces together, we have that

$$
\begin{aligned}
\Pr[|S| < 2] &= \Pr[B_{\ell_1} \notin S \cup B_{\ell_2} \notin S] \\
&\leq \Pr\left[\hat{c}_{\ell_1}(\mathbf{x}) < \frac{3n}{kC^3\sqrt{\log C}}\right] + \Pr\left[\hat{c}_{\ell_2}(\mathbf{x}) < \frac{3n}{kC^3\sqrt{\log C}}\right] \\
&\leq \beta/2 + \beta/2 \\
&= \beta
\end{aligned}
$$

which completes the proof. ◀

Proof (Theorem 8). We begin by establishing differential privacy. By Lemma 6, the functions `estimate-first-moment` and `find-interior-point` each satisfy $(\varepsilon/2, \delta/2)$-differential privacy. In the full algorithm, the output of `estimate-first-moment` is used as an input for `find-interior-point`. It follows by the standard composition lemma (see, e.g., [17]) that `interior-point-main` satisfies (ε, δ)-differential privacy.

Next we turn our attention to the probability of `interior-point-main` returning an interior point. Let P be a C-bounded distribution for some $C > 2$ and let $\mu = \mathbb{E}_{X \leftarrow P}[X]$. Finally, let $\mathbf{x} \leftarrow P^n$ where n satisfies (4). Critically, the fact that n satisfies (4) will allow for us to apply Proposition 9 and Proposition 15. By Proposition 9, we have with probability at least $1 - \beta/2$ that `estimate-first-moment(`$\mathbf{x}; \varepsilon, \delta, C$`)` returns a value $\hat{m}$ satisfying

$$
\mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|] \leq \hat{m} \leq \left(6000C\sqrt{\log C}\right) \mathop{\mathbb{E}}_{X \leftarrow P}[|X - \mu|]. \tag{9}
$$

Conditioned on (9), it follows by Proposition 15 that `find-interior-point` returns an interior point of $\mathbf{x}$ with probability at least $1 - \beta/2$. Thus, with probability at least $1 - \beta$, `interior-point-main(`$\mathbf{x}; \varepsilon, \delta, C$`)` returns an interior point of $\mathbf{x}$. ◀

4 An Algorithm for Approximate Medians

In this section, we introduce a private algorithm (Algorithm 2) for finding an α-approximate median of a distribution. We show that, if the middle 2α-percentile of the data distribution is C-bounded, then the algorithm returns an α-approximation of the median with probability at least $1 - \beta$.

As a convention in this section, we shall use P to refer to the data distribution from which $\mathbf{x}$ is sampled. We will then use P_α to refer to the middle 2α-percentile of P, that is, $P_\alpha = P \mid P \in (Q_P(0.5 - \alpha), Q_P(0.5 + \alpha))$. Note that, rather than requiring that P is C-bounded, we require that P_α is C-bounded.

Overview of Algorithm 2

Suppose we had direct sample access to the data distribution P_α. An interior point of P_α is trivially an α-approximation to the median of P. If P_α is C-bounded, then by Theorem 8, we could obtain a private α-approximation to the median. Unfortunately, we cannot assume direct sample access to P_α without infinitely-many samples. Thus, Algorithm 2 instead takes as input the dataset $\mathbf{x} \leftarrow P^n$, isolates samples $\mathbf{x_0} \subseteq \mathbf{x}$ which make up *almost* the middle 2α fraction of $\mathbf{x}$, and runs Algorithm 1 on this smaller dataset $\mathbf{x_0}$. While $\mathbf{x_0}$ is not sampled i.i.d. from P_α, we prove that, with high probability, $\mathbf{x_0}$ comes from a family of distributions similar to P_α that are C'-bounded for some $C' = O(C)$.

To construct $\mathbf{x_0}$, we isolate the middle $(2\alpha - 1/k)$-percentile of $\mathbf{x}$, for some parameter k that ends up being a function of C and α. The parameter k plays a critical role here, as it guarantees that $\mathbf{x_0}$ ends up coming from a distribution that is *contained* in P_α, rather than from a distribution that *contains* P_α. As we shall see in the analysis, this distinction allows for us to establish that the distribution P' from which $\mathbf{x_0}$ is sampled is $O(C)$-bounded.

We now introduce Theorem 18 which gives the formal guarantees of Algorithm 2:

Algorithm 2 Median Algorithm.

Function `main`$(\mathbf{x}; \varepsilon, \delta, \alpha, C)$**:**
Let $k = 1024C\alpha^{-1}$.
Let $\mathbf{x_0} = \left\{x_i \mid x_i \in \left(Q_{\mathbf{x}}\left(0.5 - \alpha + \frac{1}{2k}\right), Q_{\mathbf{x}}\left(0.5 + \alpha - \frac{1}{2k}\right)\right)\right\}$.
return `interior-point-main`$(\mathbf{x_0}; \varepsilon, \delta, 64C)$.

▶ **Theorem 18.** *Let $\beta, \varepsilon, \delta \in (0,1)$, and $\alpha \in (0, 0.25)$. Suppose P is a data distribution such that the conditional distribution on the middle 2α-percentile of P has C-bounded normalized variance for some $C > 2$. If $\mathbf{x}$ contains n datapoints where*

$$n \geq k_0 \max\left(\frac{C^3\sqrt{\log C}(\varepsilon^{-1}\ln\delta^{-1} + \ln\beta^{-1})}{\alpha}, \frac{C^2 \ln\beta^{-1}}{\alpha^2}\right)$$

for a sufficiently large positive constant k_0, then Algorithm 2 returns an α-approximation of the median with probability at least $1-\beta$. In addition, for any $\epsilon, \delta \in (0,1)$ and $C > 2$, we have that Algorithm 2 is (ϵ, δ)-differentially private.

We remark that, in the following analysis, as well as in the statement of Algorithm 2, we have not made any effort to optimize the constants involved. We have opted to give explicit constants for concreteness, but with more careful bookkeeping, these constants could almost certainly be made much smaller.

Our next lemma establishes that the endpoints of $\mathbf{x_0}$ are guaranteed to (1) be contained within the middle 2α-percentile of P; and (2) be very close to the endpoints of that middle 2α-percentile.

▶ **Lemma 19.** *Let $\beta \in (0,1)$, $\alpha \in (0, 0.25)$, and $k \geq 1$. Let P be a distribution, and let $\mathbf{x} \leftarrow P^n$ for $n \geq 108k^2\log(4/\beta)$. With probability at least $1-\beta$,*

$$Q_{\mathbf{x}}\left(0.5 - \alpha + \frac{1}{2k}\right) \in \left(Q_P(0.5-\alpha), Q_P\left(0.5 - \alpha + \frac{1}{k}\right)\right)$$

and

$$Q_{\mathbf{x}}\left(0.5 + \alpha - \frac{1}{2k}\right) \in \left(Q_P\left(0.5 + \alpha - \frac{1}{k}\right), Q_P(0.5+\alpha)\right).$$

Proof. Let $X_1, \ldots, X_n$ be n i.i.d. samples from P, and define Y_i for all $i \in \{1, \ldots, n\}$ as

$$Y_i = \begin{cases} 1 & \text{if } X_i < Q_P(0.5 - \alpha) \\ 0 & \text{otherwise.} \end{cases}$$

Define $Y \coloneqq \sum_{i=1}^{n} Y_i$. Note that $\mathbb{E}[Y] = \sum_{i=1}^{n} \Pr[X_i < Q_P(0.5-\alpha)] = (0.5-\alpha)n$. Thus,

$$\begin{aligned} \Pr\left[Q_{\mathbf{x}}\left(0.5 - \alpha + \frac{1}{2k}\right) \le Q_P(0.5-\alpha)\right] &= \Pr\left[Y \ge \left(0.5 - \alpha + \frac{1}{2k}\right) n\right] \\ &= \Pr\left[Y \ge (0.5-\alpha)n\left(1 + \frac{1}{2k(0.5-\alpha)}\right)\right] \\ &\le \Pr\left[Y \ge \mathbb{E}[Y]\left(1 + \frac{1}{k}\right)\right] \\ &\le \exp\left(\frac{-\mathbb{E}[Y]}{3k^2}\right) && \text{(by a Chernoff bound)} \\ &\le \beta/4. && \text{(by the assumption on } n\text{)} \end{aligned}$$

Likewise, for all $i \in \{1, \ldots, n\}$, let

$$Z_i = \begin{cases} 1 & \text{if } X_i > Q_P\left(0.5 - \alpha + \frac{1}{k}\right) \\ 0 & \text{otherwise.} \end{cases}$$

Let $Z \coloneqq \sum_{i=1}^{n} Z_i$. Note that $\mathbb{E}[Z] = \sum_{i=1}^{n} \Pr\left[X_i > Q_P\left(0.5 - \alpha + \frac{1}{k}\right)\right] = \left(0.5 - \alpha + \frac{1}{k}\right) n$. Thus,

$$\begin{aligned} &\Pr\left[Q_{\mathbf{x}}\left(0.5 - \alpha + \frac{1}{2k}\right) \ge Q_P\left(0.5 - \alpha + \frac{1}{k}\right)\right] = \Pr\left[Z \le \left(0.5 - \alpha + \frac{1}{2k}\right) n\right] \\ &= \Pr\left[Z \le \left(0.5 - \alpha + \frac{1}{k}\right) n\left(1 - \frac{1}{2k(0.5 - \alpha + \frac{1}{k})}\right)\right] \\ &= \Pr\left[Z \le \mathbb{E}[Z]\left(1 - \frac{1}{2k(0.5 - \alpha + \frac{1}{k})}\right)\right] \\ &\le \Pr\left[Z \le \mathbb{E}[Z]\left(1 - \frac{1}{3k}\right)\right] \\ &\le \exp\left(\frac{-\mathbb{E}[Z]}{27k^2}\right) && \text{(by a Chernoff bound)} \\ &= \exp\left(\frac{-\left(0.5 - \alpha + \frac{1}{k}\right) n}{27k^2}\right) \\ &\le \exp\left(\frac{n}{108k^2}\right) \\ &\le \beta/4. && \text{(by the assumption on } n\text{)} \end{aligned}$$

Thus, by a union bound, we have that, with probability at most $\beta/2$, $Q_{\mathbf{x}}(0.5 - \alpha + \frac{1}{2k}) \notin (Q_P(0.5-\alpha), Q_P(0.5 - \alpha + 1/k))$. By symmetry, it follows that $Q_{\mathbf{x}}(0.5 + \alpha - \frac{1}{2k}) \notin (Q_P(0.5 + \alpha - 1/k), Q_P(0.5+\alpha))$ with probability at most $\beta/2$. Thus by another union bound, the lemma holds. ◀

Our next lemma shows that, if we take a C-bounded distribution P_α and condition on being in its first $1 - 1/k$ percentile for large enough k, then the resulting conditional distribution P'_α will be $O(C)$-bounded. Note that the complement of this is not true: if the

conditional distribution P'_α is C-bounded, then the larger distribution P_α need not be $O(C)$-bounded. This is why it is critical that $\mathbf{x_0}$ is constructed in such a way that it is contained within the middle 2α-percentile of P (rather than containing the middle 2α-percentile).

▶ **Lemma 20.** *Let $C > 1$, and let P_α be a C-bounded distribution. Fix any number $k \geq 128C$. Define $P'_\alpha \coloneqq P_\alpha \mid P_\alpha \in [Q_{P_\alpha}(0), Q_{P_\alpha}(1-1/k)]$ to be P_α conditioned on being in the first $1-1/k$ percentile. It follows that P'_α is $8C$-bounded.*

Proof. We introduce notation which will be used throughout the proof. Let P''_α be the distribution P_α conditioned on being in the last $1/k$ percentile, i.e. $P''_\alpha \coloneqq P_\alpha \mid P_\alpha \in [Q_{P_\alpha}(1-1/k), Q_{P_\alpha}(1)]$. Finally, let $X \leftarrow P_\alpha$, $X' \leftarrow P'_\alpha$, and $X'' \leftarrow P''_\alpha$; let $\mu \coloneqq \mathbb{E}_{X \leftarrow P_\alpha}[X]$, $\mu' \coloneqq \mathbb{E}_{X' \leftarrow P'_\alpha}[X']$, and $\mu'' \coloneqq \mathbb{E}_{X'' \leftarrow P''_\alpha}[X'']$.

To show that P'_α is $8C$-bounded, we must show that

$$\frac{\mathbb{E}_{X'}[|X' - \mu'|^2]}{\mathbb{E}_{X'}[|X' - \mu'|]^2} \leq 8C.$$

By the C-boundedness of P_α, it suffices to show

$$\frac{\mathbb{E}_{X'}[|X' - \mu'|^2]}{\mathbb{E}_{X'}[|X' - \mu|]^2} \leq 8 \cdot \frac{\mathbb{E}_X[|X - \mu|^2]}{\mathbb{E}_X[|X - \mu|]^2}. \tag{10}$$

We break the proof into two pieces by showing

$$\mathop{\mathbb{E}}_{X'}[|X' - \mu'|^2] \leq 2 \mathop{\mathbb{E}}_{X}[|X - \mu|^2] \tag{11}$$

and

$$\mathop{\mathbb{E}}_{X'}[|X' - \mu'|]^2 \geq \frac{1}{4} \mathop{\mathbb{E}}_{X}[|X - \mu|]^2. \tag{12}$$

To prove (11), note that $\mathbb{E}[|X' - \mu'|^2] \leq \mathbb{E}[|X' - \mu|^2]$ since μ' minimizes the expectation. It follows that

$$\begin{aligned}
&\mathop{\mathbb{E}}_{X}\left[|X - \mu|^2 \cdot \mathbb{I}_{X < Q_{P_\alpha}(1-1/k)}\right] \\
&= \mathop{\mathbb{E}}_{X}\left[|X - \mu|^2 \mid X < Q_{P_\alpha}(1-1/k)\right] \cdot \Pr_X\left[X < Q_{P_\alpha}(1-1/k)\right] \\
&= \mathop{\mathbb{E}}_{X'}\left[|X' - \mu|^2\right](1-1/k)
\end{aligned}$$

which rearranges to

$$\begin{aligned}
\mathop{\mathbb{E}}_{X'}[|X' - \mu|^2] &= \frac{\mathbb{E}_X\left[|X - \mu|^2 \cdot \mathbb{I}_{X < Q_{P_\alpha}(1-1/k)}\right]}{1 - 1/k} \\
&\leq 2 \mathop{\mathbb{E}}_{X}\left[|X - \mu|^2 \cdot \mathbb{I}_{X < Q_{P_\alpha}(1-1/k)}\right] && \text{(by assumption on } k\text{)} \\
&\leq 2 \mathop{\mathbb{E}}_{X}[|X - \mu|^2],
\end{aligned}$$

and so indeed (11) is true. To prove (12), we begin by expanding $\mathbb{E}_X[|X - \mu|]$ in the following way:

$$\begin{aligned}
\mathop{\mathbb{E}}_{X}[|X - \mu|] &= \Pr_X\left[X < Q_{P_\alpha}(1-1/k)\right] \cdot \mathop{\mathbb{E}}_{X}\left[|X - \mu| \mid X < Q_{P_\alpha}(1-1/k)\right] \\
&\qquad + \Pr_X\left[X \geq Q_{P_\alpha}(1-1/k)\right] \cdot \mathop{\mathbb{E}}_{X}\left[|X - \mu| \mid X \geq Q_{P_\alpha}(1-1/k)\right] \\
&= \frac{k-1}{k} \cdot \mathop{\mathbb{E}}_{X'}[|X' - \mu|] + \frac{1}{k} \cdot \mathop{\mathbb{E}}_{X''}[|X'' - \mu|] \\
&\leq \frac{k-1}{k} \cdot \mathop{\mathbb{E}}_{X'}[|X' - \mu'|] + \frac{k-1}{k} \cdot |\mu' - \mu| + \frac{1}{k} \cdot \mathop{\mathbb{E}}_{X''}[|X'' - \mu|]
\end{aligned}$$

which rearranges to

$$\mathop{\mathbb{E}}_{X'}[|X' - \mu'|] \geq \frac{k}{k-1}\left(\mathop{\mathbb{E}}_{X}[|X-\mu|] - \frac{k-1}{k}|\mu' - \mu| - \frac{1}{k}\mathop{\mathbb{E}}_{X''}[|X''-\mu|]\right). \tag{13}$$

To lowerbound $\mathbb{E}_{X'}[|X' - \mu'|]$ as in (12), we seek to upperbound $|\mu' - \mu|$ and $\mathbb{E}[|X'' - \mu|]$ in terms of $\mathbb{E}_X[|X - \mu|]$. As an intermediate step, we can express $|\mu' - \mu|$ and $\mathbb{E}[|X'' - \mu|]$ in terms of $(\mu'' - \mu)$ and then upperbound $(\mu'' - \mu)$ in terms of $\mathbb{E}_X[|X - \mu|]$. It is not difficult to show that $|\mu - \mu'| = \frac{1}{k-1}(\mu'' - \mu)$ and $\mathbb{E}[|X'' - \mu|] = \mu'' - \mu$ using both Chebyshev's and the C-boundedness of P_α (the proofs are deferred to the full version of the paper).

The most technically challenging piece is upperbounding $\mu'' - \mu$. We introduce the following claim which gives an upperbound on $\mu'' - \mu$ (the proof is deferred to the full version of the paper):

▷ Claim 21. If $k \geq 128C$, then $\mu'' - \mu \leq 3\sqrt{Ck}\,\mathbb{E}_X[|X - \mu|]$.

Putting the pieces together, we see that

$$\begin{aligned}
\mathop{\mathbb{E}}_{X'}[|X' - \mu'|] &\geq \frac{k}{k-1}\left(\mathop{\mathbb{E}}_{X}[|X-\mu|] - \frac{k-1}{k}|\mu' - \mu| - \frac{1}{k}\mathop{\mathbb{E}}_{X''}[|X''-\mu|]\right) \\
&= \frac{k}{k-1}\left(\mathop{\mathbb{E}}_{X}[|X-\mu|] - \frac{1}{k}(\mu'' - \mu) - \frac{1}{k}(\mu'' - \mu)\right) \\
&\geq \frac{k}{k-1}\left(1 - \frac{2}{k}\cdot 3\sqrt{Ck}\right)\mathop{\mathbb{E}}_{X}[|X-\mu|] && \text{(by Claim 21)} \\
&\geq \frac{1}{2}\mathop{\mathbb{E}}_{X}[|X-\mu|]. && \text{(by assumption on } k\text{)}
\end{aligned}$$

This implies $\mathbb{E}_{X'}[|X' - \mu'|]^2 \geq \frac{1}{4}\mathbb{E}_X[|X - \mu|]^2$, proving (12) and thus completing the proof of the lemma. ◀

Applying Lemma 20 twice, one arrives at the two-sided version of it that we use in the proof of the theorem:

▶ **Lemma 22.** *Let $C > 1$, and let P_α be a C-bounded distribution. Define $P_m \coloneqq P_\alpha \mid P_\alpha \in [Q_{P_\alpha}(1/k_1), Q_{P_\alpha}(1 - 1/k_2)]$ for some k_1, k_2. If $k_1, k_2 \geq 2048C$, then P_m is $64C$-bounded.*

Proof. Define $P'_\alpha = P_\alpha \mid P_\alpha \in [Q_{P_\alpha}(0), Q_{P_\alpha}(1 - 1/k_2)]$. As P_α is C-bounded and $k_2 \geq 128C$, we have by Lemma 20 that P'_α is $8C$-bounded. Next note that $P_m = P_\alpha \mid P_\alpha \in [Q_{P_\alpha}(1/k_1), Q_{P_\alpha}(1 - 1/k_2)] = P'_\alpha \mid P'_\alpha \in \left[Q_{P_\alpha}\left(\frac{1}{k_1(1-1/k_2)}\right), Q_{P_\alpha}(1)\right]$. As P'_α is C'-bounded and $k_1(1 - 1/k_2) \geq k_1/2 \geq 128C'$, we have by Lemma 20 that P_m is $8C'$-bounded and thus P_m is $64C$-bounded. ◀

Proof (Theorem 18). We begin by establishing differential privacy. Suppose $\mathbf{x}$ and $\mathbf{x}'$ differ only in one data point. Note that to obtain $\mathbf{x_0}$, we sort the elements in $\mathbf{x}$ and take all the elements that have ranks between $\lfloor n \cdot (0.5 - \alpha + \frac{1}{2k})\rfloor$ and $\lfloor n \cdot (0.5 + \alpha - \frac{1}{2k})\rfloor$. It is straightforward to show that if we change one data point in $\mathbf{x}$, at most one data point in $\mathbf{x_0}$ will be changed. Moreover, previously in Theorem 8, we have shown that the procedure for finding the interior point is (ϵ, δ)-differentially private. Hence, Algorithm 2 is (ϵ, δ)-differentially private.

Now we analyze the accuracy guarantee of Algorithm 2. By Lemma 19, we have that with probability at least $1 - \beta/2$ that

$$Q_{\mathbf{x}}\left(0.5 - \alpha + \frac{1}{2k}\right) \in \left(Q_P(0.5 - \alpha), Q_P\left(0.5 - \alpha + \frac{1}{k}\right)\right) \tag{14}$$

and

$$Q_{\mathbf{x}}\left(0.5+\alpha-\frac{1}{2k}\right) \in \left(Q_P\left(0.5+\alpha-\frac{1}{k}\right), Q_P(0.5+\alpha)\right) \tag{15}$$

For the rest of the proof, we condition on some arbitrary fixed outcome for the values of $Q_{\mathbf{x}}(0.5-\alpha+\frac{1}{2k})$ and $Q_{\mathbf{x}}(0.5+\alpha-\frac{1}{2k})$ such that (14) and (15) are satisfied.

Note that, once we condition on the outcomes of $Q_{\mathbf{x}}(0.5-\alpha+\frac{1}{2k})$ and $Q_{\mathbf{x}}(0.5+\alpha-\frac{1}{2k})$, then $\mathbf{x_0}$ consists of i.i.d. samples from the distribution $P_m \coloneqq P \mid P \in (Q_x(0.5-\alpha+\frac{1}{2k}), Q_x(0.5+\alpha-\frac{1}{2k}))$. By (14) and (15), this distribution P_m can be expressed as

$$P_m = P|P \in \left(Q_P\left(0.5-\alpha+\frac{1}{k_1}\right), Q_P\left(0.5+\alpha-\frac{1}{k_2}\right)\right)$$

for some $k_1, k_2 \geq k$.

Finally, since $k = 2048C/(2\alpha)$, we can apply Lemma 22, which says that P_m is $64C$-bounded. Also note that, by assumption on n,

$$|\mathbf{x_0}| = \left(2\alpha-\frac{1}{k}\right)n > k_0 C^3\sqrt{\log C}\cdot\left(\varepsilon^{-1}\ln\delta^{-1}+\ln\beta^{-1}\right).$$

Thus, we can apply Theorem 8, which says that, with probability at least $1-\beta/2$, the return value of `interior-point-main(`$\mathbf{x_0}$$;\varepsilon,\delta,64C)$ will be an interior point to $\mathbf{x_0}$. This, in turn, is an α-approximation of the median of P with probability at least $1-\beta$. ◀

References

1. Maryam Aliakbarpour, Rose Silver, Thomas Steinke, and Jonathan Ullman. Differentially private medians and interior points for non-pathological data. *arXiv preprint*, 2023. `arXiv:2305.13440`.
2. Noga Alon, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private pac learning implies finite littlestone dimension. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 852–860, 2019. `arXiv:1806.00949`.
3. Hilal Asi and John C Duchi. Near instance-optimality in differential privacy. *arXiv preprint*, 2020. `arXiv:2005.10630`.
4. Amos Beimel, Shay Moran, Kobbi Nissim, and Uri Stemmer. Private center points and learning of halfspaces. In *Conference on Learning Theory*, pages 269–282. PMLR, 2019.
5. Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, RANDOM-APPROX '13, pages 363–378. Springer, 2013.
6. Daniel Berend and Aryeh Kontorovich. A sharp estimate of the binomial mean absolute deviation with applications. *Statistics & Probability Letters*, 83(4):1254–1259, 2013. `doi:10.1016/j.spl.2013.01.023`.
7. Victor-Emmanuel Brunel and Marco Avella-Medina. Propose, test, release: Differentially private estimation with high probability. *arXiv preprint*, 2020. `arXiv:2002.08774`.
8. Mark Bun, Cynthia Dwork, Guy N Rothblum, and Thomas Steinke. Composable and versatile privacy via truncated cdp. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 74–86, 2018.
9. Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. Differentially private release and learning of threshold functions. In *IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 634–649, 2015.

10 Edith Cohen, Xin Lyu, Jelani Nelson, Tamás Sarlós, and Uri Stemmer. Optimal differentially private learning of thresholds and quasi-concave optimization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 472–482, 2023.

11 Edwin L. Crow. The mean deviation of the poisson distribution. *Biometrika*, 45(3/4):556–559, 1958. URL: `http://www.jstor.org/stable/2333201`.

12 Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, STOC '09, pages 371–380. ACM, 2009.

13 Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Conference on Theory of Cryptography*, TCC '06, pages 265–284, New York, NY, USA, 2006.

14 Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis of online and differentially private learning. *Advances in Neural Information Processing Systems*, 33:9203–9215, 2020.

15 Haim Kaplan, Katrina Ligett, Yishay Mansour, Moni Naor, and Uri Stemmer. Privately learning thresholds: Closing the exponential gap. In *Conference on Learning Theory*, pages 2263–2285. PMLR, 2020.

16 Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz – Zentrum für Informatik, 2018. `arXiv:1711.03908`.

17 Adam Smith and Jonathan Ullman. Privacy in statistics and machine learning spring 2021 lecture 9: Approximate differential privacy. URL: `https://dpcourse.github.io/2021-spring/index.html`.

18 Christos Tzamos, Emmanouil-Vasileios Vlatakis-Gkaragkounis, and Ilias Zadik. Optimal private median estimation under minimal distributional assumptions. *Advances in Neural Information Processing Systems*, 33:3301–3311, 2020.

19 Salil Vadhan. The complexity of differential privacy, 2016. URL: `https://privacytools.seas.harvard.edu/files/privacytools/files/complexityprivacy_1.pdf`.

A Appendix

In Table 1 we give some examples of distributions with C-bounded normalized variance, with explicit values of C. Lemmas 23 and 24 show that combinations of these distributions also have bounded normalized variance.

Table 1 Examples of C for various natural families of distributions.

Distribution	**PDF**	C	
Uniform distribution over $\mathcal{U}[a,b]$:	$p(x) = \frac{1}{b-a}$	$\frac{4}{3}$	
Normal distribution: $\mathcal{N}(\mu,\sigma)$	$p(x) = e^{(x-\mu)^2/(2\sigma^2)} / \left(\sigma\sqrt{2\,\pi}\right)$	$\frac{\pi}{2}$	
Exponential distribution Exp[λ]	$p(x) = \lambda e^{-\lambda\, x}$	$\frac{e^2}{4}$	
Laplace distribution Lap[μ,β]	$p(x) = e^{-\frac{\lvert x-\mu\rvert}{\beta}}/(2\beta)$	2	
Binomial distribution Bin(n,q), $1 \le nq \le n-1$	$p(k) = \binom{n}{k} q^k (1-q)^{n-k}$	2	[6]
Poisson distribution Pois(λ), $\lambda \ge 1$	$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$	14	[11]

▶ **Lemma 23.** *Let X and Y be independent random variables. Assume X has C_1-bounded normalized variance and Y has C_2-bounded normalized variance. Then $X+Y$ has (C_1+C_2)-bounded normalized variance.*

Proof. Without loss of generality, we may assume $\mathbb{E}[X] = \mathbb{E}[Y] = 0$. Thus

$$\mathbb{E}[(X+Y)^2] = \mathbb{E}[X^2 + 2XY + Y^2] = \mathbb{E}[X^2] + \mathbb{E}[Y^2] \leq C_1 \mathbb{E}[|X|]^2 + C_2 \mathbb{E}[|Y|]^2.$$

By Jensen's inequality,

$$\mathbb{E}[|X+Y|] = \mathop{\mathbb{E}}_{X}[\mathop{\mathbb{E}}_{Y}[|X+Y|]] \geq \mathop{\mathbb{E}}_{X}[|\mathop{\mathbb{E}}_{Y}[X-Y]|] = \mathbb{E}[|X|].$$

Similarly, $\mathbb{E}[|X+Y|] \geq \mathbb{E}[|Y|]$. Now

$$\frac{\mathbb{E}[(X+Y)^2]}{\mathbb{E}[|X+Y|]^2} \leq \frac{C_1 \mathbb{E}[|X|]^2 + C_2 \mathbb{E}[|Y|]^2}{\max\{\mathbb{E}[|X|]^2, \mathbb{E}[|Y|]^2\}} = \min\left\{C_1 + C_2 \frac{\mathbb{E}[|Y|]^2}{\mathbb{E}[|X|]^2}, C_1 \frac{\mathbb{E}[|X|]^2}{\mathbb{E}[|Y|]^2} + C_2\right\},$$

which is at most $C_1 + C_2$ as required. ◀

▶ **Lemma 24.** *Let X be a random variable that has C-bounded normalized variance. Let $a, b \in \mathbb{R}$. Then $aX+b$ has C-bounded normalized variance.*

Proof. Let $\mu = \mathbb{E}[X]$. Let $\mu' = \mathbb{E}[aX+b] = a\mu + b$. Then

$$\mathbb{E}[(aX+b-\mu')^2] = \mathbb{E}[(a(X-\mu)+b-b)^2] = \mathbb{E}[a^2(X-\mu)^2] = a^2 \mathbb{E}[(X-\mu)^2]$$

and

$$\mathbb{E}[|aX+b-\mu'|]^2 = \mathbb{E}[|a(X-\mu)+b-b|]^2 = \mathbb{E}[a|X-\mu|]^2 = a^2 \mathbb{E}[|X-\mu|]^2.$$

So

$$\frac{\mathbb{E}[(aX+b-\mu')^2]}{\mathbb{E}[|aX+b-\mu'|]^2} = \frac{a^2 \mathbb{E}[(X-\mu)^2]}{a^2 \mathbb{E}[|X-\mu|]^2} = \frac{\mathbb{E}[(X-\mu)^2]}{\mathbb{E}[|X-\mu|]^2} \leq C,$$

as required. ◀