Universal Private Estimators

Wei Dong
Hong Kong University of Science and Technology
Hong Kong, China
wdongac@cse.ust.hk

Ke Yi
Hong Kong University of Science and Technology
Hong Kong, China
yike@cse.ust.hk

ABSTRACT

We present *universal* estimators for the statistical mean, variance, and scale (in particular, the interquartile range) under pure differential privacy. These estimators are universal in the sense that they work on an arbitrary, unknown continuous distribution $\mathcal{P}$ over $\mathbb{R}$, while yielding strong utility guarantees except for ill-behaved $\mathcal{P}$. For certain distribution families like Gaussians or heavy-tailed distributions, we show that our universal estimators match or improve existing estimators, which are often specifically designed for the given family and under *a priori* boundedness assumptions on the mean and variance of $\mathcal{P}$. This is the first time these boundedness assumptions are removed under pure differential privacy. The main technical tools in our development are instance-optimal empirical estimators for the mean and quantiles over the unbounded integer domain, which can be of independent interest.

CCS CONCEPTS

• **Security and privacy**; • **Mathematics of computing** → **Probability and statistics**; • **Theory of computation** → **Data structures design and analysis**;

KEYWORDS

Differential privacy; Statistical Estimation; Free assumption

ACM Reference Format:
Wei Dong and Ke Yi. 2023. Universal Private Estimators. In *Proceedings of the 42nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS '23), June 18–23, 2023, Seattle, WA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3584372.3588669

1 INTRODUCTION

Parameter estimation is a central problem in statistics, data mining, and machine learning. Let $\mathcal{P}$ be a continuous probability distribution over $\mathbb{R}$ with density function (pdf) $f(x)$, and let $F(x)$ be its cumulative distribution function (CDF). We consider the following three fundamental parameters, mean, variance, and IQR:

$$\mu_{\mathcal{P}} = \int_{-\infty}^{\infty} x f(x), \mathrm{d}x, \quad \sigma^2_{\mathcal{P}} = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,\mathrm{d}x,$$
$$\mathrm{IQR}_{\mathcal{P}} = F^{-1}(3/4) - F^{-1}(1/4).$$

PODS '23, June 18–23, 2023, Seattle, WA, USA

ACM ISBN 979-8-4007-0127-6/23/06…$15.00
https://doi.org/10.1145/3584372.3588669

Note that the *interquartile range (IQR)* is a widely used parameter for the *scale* of $\mathcal{P}$, but the particular choices of $1/4$ and $3/4$ are not very important: changing them to other constants does not affect our results (for both error bound and the requirement of n) asymptotically. For simplicity, we omit the subscript $\mathcal{P}$ when there is no confusion.

Given an i.i.d. sample $D = (X_1, \ldots, X_n)$ drawn from $\mathcal{P}^n$, the standard estimators for these parameters are (we reorder D such that $X_1 \le \cdots \le X_n$):

$$\mu(D) = \frac{1}{n}\sum X_i, \mathrm{d}x, \quad \sigma^2(D) = \frac{1}{n}\sum (X_i - \mu(D))^2,$$
$$\mathrm{IQR}(D) = X_{3n/4} - X_{n/4},$$

which are often called the *sample* or *empirical* mean, variance, and IQR. They all converge to the true parameter respectively at a rate of $O(1/\sqrt{n})$, and the difference between the empirical parameter and the statistical parameter is referred to as the *sampling error*. Importantly, all these estimators are *universal*, namely, they work on an arbitrary, unknown $\mathcal{P}$. The $O(1/\sqrt{n})$ convergence rate is optimal for many families of distributions, but not all. For instance, the mid-range estimator $(X_1 + X_n)/2$ is a better estimator of μ for uniform distributions with a convergence rate of $O(1/n)$. However, such distribution-specific estimators are less used in practice as we usually do not know which family $\mathcal{P}$ is chosen from, and they may fail miserably when the distributional assumption does not hold (e.g., the mid-range estimator is a very bad estimator of the Gaussian mean).

In this paper, we design universal estimators under *differential privacy (DP)* [33]. A randomized mechanism $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ satisfies (ε, δ)-DP if for any two neighboring datasets $D \sim D'$ (i.e., D and D' differ by one record), and any $\mathcal{S} \subseteq \mathcal{Y}$,

$$\Pr[\mathcal{M}(D) \in \mathcal{S}] \le e^{\varepsilon} \cdot \Pr[\mathcal{M}(D') \in \mathcal{S}] + \delta, \tag{1}$$

for some privacy parameters $0 < \varepsilon < 1, 0 \le \delta < 1/n^{\omega(1)}$. For statistical estimation problems, the high-privacy regime (e.g., $\varepsilon < 1/\sqrt{n}$) is more interesting; otherwise, the error would be dominated by the sampling error for many distributions (i.e., privacy is free). This is because the privacy error is $\tilde{O}(1/(\epsilon n))$ while the sampling error is $\tilde{O}(1/\sqrt{n})$. The case $\delta = 0$ is often called *pure DP*, abbreviated as ε-DP. It is preferable than the $\delta > 0$ case, since δ corresponds to the probability of catastrophic privacy breaches. However, there are strong separation results showing that for certain problems, ε-DP is strictly harder to achieve than (ε, δ)-DP [10, 16, 21, 36, 62]. Note that, when designing a private estimator, the DP guarantee should hold for any two neighboring datasets D, D', and (1) is only over the internal randomness of $\mathcal{M}$. When analyzing its utility, however, the randomness in both D and $\mathcal{M}$ is taken into consideration.

In the past several years, quite a number of private estimators have been proposed in the literature as summarized in Table 1. With

	μ	σ^2	IQR
ε-DP	A1, A2, A3 [58] A1, A2, A3 [45] A1, A2, A3 [41] A1, A2, A3 [14] A1, A2, A3 [11] A1, A2 [44] A1, A2 [37]	A1, A2, A3 [45] A2, A3 [41] A1, A2, A3 [14] A2, A3 [11]	None
(ε, δ)-DP	A3 [45] A1, A2 [17] A1, A2, A3 [41] A2, A3 [14] A1, A2, A3 [18] A1, A2, A3 [11] A3 [1] A1, A2, A3 [38] A3 [42] A3 [12] A3 [49] A3 [6] A3 [46]	A3 [45] A2, A3 [41] A2, A3 [11] A3 [1] A3 [42] A3 [49] A3 [6] A3 [46]	[30]

Table 1: Summary of existing private estimators[1] and their assumptions.

the exception of the IQR estimator of [30], which only satisfies (ε, δ)-DP, none of them is universal. They all rely on the following three assumptions or a subset of them:

- A1. a predefined range for the mean, i.e., $\mu \in [-R, R]$;
- A2. a predefined range for the variance, i.e., $\sigma^2 \in [\sigma^2_{\min}, \sigma^2_{\max}]$, as well as ranges for the higher moments if applicable;
- A3. $\mathcal{P}$ is chosen from a specific family of distributions such as Gaussian.

In particular, their reliance on A1/A2 is both algorithmic and analytical, i.e., these estimators need $R, \sigma_{\min}, \sigma_{\max}$ together with D as the input, and the utility guarantees also depend on these *a priori* bounds. The reliance on A3 is only analytical; when we write A3 in Table 1, the corresponding estimator does not offer utility guarantees when $\mathcal{P}$ is chosen outside the specified family.[2]

In this paper, we design universal private estimators under pure DP for μ, σ^2, and IQR without these assumptions while achieving the same or better utilities. As shown in Table 1, this is the first time A1/A2 have been removed under pure-DP. Under (ε, δ)-DP, a number of prior works [1, 6, 12, 14, 42, 45, 46, 49] show how A1/A2 can be removed, using *stability* based techniques [13, 15, 31, 60, 62], the *propose-test-release* framework [30], or the *truncated distribution* [19]. However, these techniques fundamentally do not work under pure DP. More precisely, for the stability based techniques and the truncated distribution, even the output domain is different for neighboring datasets. The propose-test-release framework by nature must have a small probability that the privacy is breached, thus can only achieve (ϵ, δ)-DP.

As a necessary consequence, the utility guarantees of our estimators depend on the properties of the unknown $\mathcal{P}$, namely, they yield instance-specific utility bounds. As we shall see below, compared to existing estimators that aim at optimizing the worst case (i.e., minimax bounds), our instance-specific bounds are no worse except for an ill-behaved $\mathcal{P}$, while could be much better for most realistic $\mathcal{P}$'s. Finally, all our estimators can be implemented efficiently in $O(n \log n)$ time.

Our general approach is as follows. We first study the empirical problem, in particular, estimating the empirical mean $\mu(D)$ and the τ-th quantile X_τ for any given D. These empirical estimators only work over discrete domains, but we can apply them in the statistical setting by appropriately discretizing $\mathbb{R}$. To remove A1/A2, we make our empirical estimators work over an infinite but discrete domain, namely, $\mathbb{Z}$. To remove A3, we show that the errors achieved by our empirical estimators are instance-optimal, hence adaptive to an arbitrary $\mathcal{P}$ when applied in the statistical setting. Although our main motivation is in the statistical setting, the instance-optimality of our empirical estimators is of independent interest.

[2] [49] can handle different distribution families but need to manually adjust the algorithm based on the distribution family.

1.1 Empirical Estimators

Let $D = \{X_1, \ldots, X_n\}$ be a multiset drawn from $\mathbb{Z}$, and assume $X_1 \le \cdots \le X_n$. Estimating $\mu(D)$ and X_τ under DP has been studied previously, but existing algorithms either do not provide utility guarantees [3, 4, 50, 55] or only work over a finite domain $[N] = \{0, 1, \ldots, N\}$ [7, 38, 53].

To reduce the domain from $\mathbb{Z}$ to a finite one, the natural idea is to use the *empirical range* $\mathcal{R}(D) = [X_1, X_n]$ as the domain. However, doing so violates DP, and we must use a privatized $\tilde{\mathcal{R}}(D)$. A good $\tilde{\mathcal{R}}(D)$ should be close to $\mathcal{R}(D)$ in both location and scale. We thus approach the problem in two steps. First, we obtain a privatized *radius* of D, which is defined as $\mathrm{rad}(D) = \max_i |X_i|$. We show that our privatized radius is not too much larger than $\mathrm{rad}(D)$ while covering all but $O(\log\log(\mathrm{rad}(D))/\varepsilon)$ elements[3] in D:

Theorem 1.1 (Theorem 2.1, informal). *There exists an ε-DP mechanism such that for any $D \in \mathbb{Z}^n$, it returns a $\widetilde{\mathrm{rad}}(D)$ such that*[4] $\widetilde{\mathrm{rad}}(D) \le 2 \cdot \mathrm{rad}(D)$ *and* $\left|D \cap \overline{\left[-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)\right]}\right| = O\left(\frac{1}{\varepsilon} \log\log\left(\mathrm{rad}(D)\right)\right)$.

In the second step, we try to find a rough location of $\mathcal{R}(D)$. As we have bounded most elements into $\left[-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)\right]$, this can be done by using a finite-domain private median (Appendix B.5). Then we shift D to the median and invoke again our private radius estimator. This results in a privatized $\tilde{\mathcal{R}}(D)$, whose width is not too much larger than the actual width $\gamma(D) = X_n - X_1$:

Theorem 1.2 (Theorem 2.2, informal). *There exists an ε-DP mechanism such that for any $D \in \mathbb{Z}^n$ and n not too small, it returns a range $\tilde{\mathcal{R}}(D)$ such that $|\tilde{\mathcal{R}}(D)| \le 4 \cdot \gamma(D)$, and* $\left|D \cap \overline{\tilde{\mathcal{R}}(D)}\right| = O\left(\frac{1}{\varepsilon} \log\log\left(\gamma(D)\right)\right)$.

We can now invoke existing finite-domain empirical mean estimators [7, 38, 53] using $\tilde{\mathcal{R}}(D)$ as the domain, but it turns out that using $\tilde{\mathcal{R}}(D)$ directly with the *clipped mean estimator* (Appendix B.6) yields an even better result:

Theorem 1.3 (Theorem 2.3, informal). *There exists an ε-DP mechanism such that for any $D \in \mathbb{Z}^n$ and n not too small, it returns a $\tilde{\mu}(D)$ such that* $|\tilde{\mu}(D) - \mu(D)| = O\left(\frac{\gamma(D)}{\varepsilon n} \log\log\left(\gamma(D)\right)\right)$.

[3] In this paper, we use e as the base of log and define $\log(x) = 1$ for any $x \le e$, unless stated otherwise.

[4] All results stated in Section 1 hold with constant success probability.

[38, 62] show that the width $\gamma(D)$ is an instance-specific lower bound. More precisely, any mean estimator under DP (pure or not) has to incur an error of $\Omega(\gamma(D)/n)$ on D or one of its *in-neighbors* (see Appendix B.3 for more details), so a result like Theorem 1.3 can be considered instance-optimal, where the extra $O(\log\log(\gamma(D))/\varepsilon)$ factor is the *optimality ratio*. In contrast, the optimality ratio in [38] is $O(\log N/\varepsilon)$[5]. Thus, we obtain an improvement even in the finite-domain case. Furthermore, we show that the optimality ratio cannot be better than $O(\log\log N/\varepsilon)$ in the finite-domain case:

Theorem 1.4 (Theorem 2.4). *For any ε, any integer $N \geq 1$, $n > \log\log_2 N/\varepsilon$, and any ε-DP mechanism $\mathcal{M} : [N]^n \to \mathbb{R}$, there exists $D \in [N]^n$, such that $|\mathcal{M}(D) - \mu(D)| \geq \frac{\gamma(D)}{3\varepsilon n}\log\log_2(N)$.*

For quantile estimation, there exists a finite-domain estimator (Appendix B.5) that achieves a rank error of $O(\log N/\varepsilon)$. Invoking it with $\tilde{\mathcal{R}}(D)$ immediately yields:

Theorem 1.5 (Theorem 2.5, informal). *There exists an ε-DP mechanism such that for any $D \in \mathbb{Z}^n$, any $1 \leq \tau \leq n$, and n not too small, it returns a value $\tilde{X}_\tau$ such that $X_{\tau-t} \leq \tilde{X}_\tau \leq X_{\tau+t}$,[6] for some $t = O\left(\frac{1}{\varepsilon}\log(\gamma(D))\right)$.*

In the finite-domain case, it is known that the rank error has to be $\Omega(\log N/\varepsilon)$ for at least one D, by a reduction from the *interior point problem* [9, 16]. In contrast, our error guarantee is a more instance-specific one, which is also worst-case optimal in the finite-domain case.

In addition, it is worth pointing out that sum estimation is equivalent to answering self-join-free aggregation queries in a relational database under user-level privacy protection [22], which has been widely researched in database community. In that problem, the state-of-the-art algorithm [22] achieves the error $O(\frac{\mathrm{rad}(D)}{\varepsilon}\log(N)\log\log(N))$ and also requires a domain assumption N. Consequently, our result also yields a significant in that problem.

1.2 Statistical Mean Estimation

Next, we move onto the statistical setting, where D is an i.i.d. sample drawn from some arbitrary, unknown $\mathcal{P}$. Before we can apply our infinite-domain empirical mean estimator (Theorem 1.3), we have to discretize $\mathbb{R}$. Since the sampling error is already $O(\sigma/\sqrt{n})$, a bucket size of $b \leq \sigma/n$ would suffice. However, σ is not known; actually, estimating σ is another mean estimation problem. Under assumption A2, prior work [17, 41, 43, 45] simply used $b = \sigma_{\min}/n$ as the bucket size. Without any assumptions, we seek to find a privatized lower bound of σ and use that as the bucket size. After that, we can apply Theorem 1.3, but this leads to sub-optimal errors in the statistical setting. The reason is that in the empirical setting, we wish to minimize the number of points outside $\tilde{\mathcal{R}}(D)$, which translates into the optimality ratio. When D is an i.i.d. sample, the points in D are more well-behaved and we can use a smaller $\tilde{\mathcal{R}}(D)$ to clip D more aggressively. Our idea is thus to find $\tilde{\mathcal{R}}(D')$ on a sub-sample D' of D and apply the clipped mean estimator. It turns out $|D'| = \varepsilon n$ is the right sub-sample size, which yields our main result on a universal private mean estimator:

Theorem 1.6 (Theorem 3.5, informal). *There exists an ε-DP mechanism such that for any $\mathcal{P}$, given $D \sim \mathcal{P}^n$, if*

$$n > \Omega\left(\frac{1}{\varepsilon}\log\log\frac{1}{\varphi(1/16)} + \frac{1}{\varepsilon}\log\log(\mathrm{IQR}) + \frac{1}{\varepsilon}\log\frac{|\mu|+\sigma+\gamma(\varepsilon n)}{\varphi(1/16)}\right), \tag{2}$$

then it returns a $\tilde{\mu}$ such that

$$|\mu-\tilde{\mu}| = O\Big(\min_{\xi\geq 10\cdot\gamma(\varepsilon n)+2\sigma}\Big(\Big|\mathsf{E}\big[X<\mu-\xi\big]+\mathsf{E}\big[X>\mu+\xi\big]\Big| + \frac{\xi}{\varepsilon n}\log\log\frac{\gamma(\varepsilon n)}{\varphi(1/16)}\Big)+\frac{\sigma}{\sqrt{n}}\Big). \tag{3}$$

The formal definitions of $\varphi(1/16)$, $\gamma(\varepsilon n)$, $\mathsf{E}[X<\mu-\xi]$, and $\mathsf{E}[X>\mu+\xi]$ are given in Appendix B.1. Roughly speaking, $\varphi(1/16)$ is the minimum width of any interval with a probability mass $1/16$, which is strictly positive for any continuous distribution $\mathcal{P}$. This term is required due to the searching for a proper bucket size. For all well-behaved $\mathcal{P}$, $\varphi(1/16) = \Theta(\sigma)$, but it may get arbitrarily small (e.g., when f has a very narrow and high peak), which we call ill-behaved. Nevertheless, we would like to stress that (1) our algorithm does not need to know $\varphi(1/16)$ *a priori* (the analysis needs it *a posteriori*); (2) our dependency on $1/\varphi(1/16)$ will be logarithmic or even log log; and (3) we did not try to optimize the constant $1/16$. $\gamma(\varepsilon n)$ is a constant-probability bound on $\gamma(D') = X'_{\varepsilon n} - X'_1$ when D' is a random sample of size εn drawn from $\mathcal{P}$, while $\mathsf{E}[X<\mu-\xi]$ and $\mathsf{E}[X>\mu+\xi]$ are the contributions to μ from the regions outside $[\mu-\xi,\mu+\xi]$, which correspond to (part of) the bias in $\tilde{\mu}$. The last term in the $\min\{\dots\}$ of (3) is the DP noise (both bias and variance). Importantly, the achieved error is the best bias-variance trade-off over all possible $\xi > 10\cdot\gamma(\varepsilon)+2\sigma$. The last term in (3) is the sampling error, which exists even in the non-private setting, so it does not depend on ε.

Most prior works in the statistical setting state their results in terms of sample complexity, namely, what is the required sample size n for achieving error α. Our lower bound requirement (2) on n easily translates into a term in the sample complexity, but it is cumbersome to rewrite (3) due to the use of $\gamma(\varepsilon n)$ and the $\min_\xi$. To facilitate the comparison, below we relax $\gamma(\varepsilon n)$ appropriately and consider some fixed ξ. This will result in simpler (but possibly looser) versions of Theorem 1.6 in terms of the sample complexity. We may also use the $\tilde{O}$ notation to suppress polylogarithmic factors in n, $\frac{1}{\alpha}$, $\log|\mu|$, $\log\sigma$, $\log\frac{1}{\varphi(1/16)}$, $\log R$, $\log\frac{\sigma_{\max}}{\sigma_{\min}}$.

Gaussian distributions. If $\mathcal{P}$ is a Gaussian, then $\varphi(1/16) = \Theta(\sigma)$ and $\gamma(\varepsilon n) = \tilde{O}\left(\sigma\sqrt{\log(\varepsilon n)}\right)$. We fix $\xi = c\cdot\sigma\sqrt{\log(\varepsilon n)}$ for some large constant c. Then Theorem 1.6 simplifies into:

Theorem 1.7 (Theorem 3.6). *For any Gaussian $\mathcal{P}$ and any $\alpha > 0$, the ε-DP mechanism from Theorem 1.6 takes $n = \tilde{O}\left(\frac{1}{\varepsilon}\log\frac{|\mu|}{\sigma} + \frac{\sigma^2}{\alpha^2} + \frac{\sigma}{\varepsilon\alpha}\right)$ samples and returns a $\tilde{\mu}$ such that $|\tilde{\mu}-\mu| \leq \alpha$.*

For Gaussian mean, [45] and [11, 41] gave two ε-DP mechanisms under A1/A2. Their sample complexities are $n = \tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma_{\min}} + \frac{\sigma^2}{\alpha^2} + \frac{\sigma}{\varepsilon\alpha}\right)$ and $n = \tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma} + \frac{1}{\varepsilon}\log\frac{\sigma_{\max}}{\sigma_{\min}} + \frac{\sigma^2}{\alpha^2} + \frac{\sigma}{\varepsilon\alpha}\right)$, respectively, both inferior to Theorem 1.7.

[5] The optimality ratio stated in [38] is $O(\sqrt{\log N/\rho})$, which holds under ρ-CDP; for pure DP, it is $O(\log N/\varepsilon)$

[6] Define $X_i = X_n$ for $i > n$ and $X_i = X_1$ for $i < 1$.

[45] show that $\Omega\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma}+\frac{\sigma^2}{\alpha^2}+\frac{\sigma}{\varepsilon\alpha}\right)$ samples are necessary. In fact, what they have proved is a worst-case lower bound, i.e., for any R, σ, and any ε-DP mechanism $\mathcal{M}$, there exists a Gaussian distribution $\mathcal{P}$ with $|\mu_{\mathcal{P}}|\le R, \sigma_{\mathcal{P}}=\sigma$ such that $\mathcal{M}$ requires $\Omega\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma}+\frac{\sigma^2}{\alpha^2}+\frac{\sigma}{\varepsilon\alpha}\right)$ samples to estimate $\mu_{\mathcal{P}}$ within an error of α. Our mechanism indeed requires this many samples on a $\mathcal{P}$ with $|\mu_{\mathcal{P}}|=R, \sigma_{\mathcal{P}}=\sigma$, thus no contradiction.

Heavy-tailed distributions. Next, we consider the case where $\mathcal{P}$ has a finite kth central moment μ_k for some $k\ge 2$. In this case, we have $\gamma(\varepsilon n) < O\left((\varepsilon n\mu_k)^{1/k}\right)$. Fixing $\xi = c\cdot(\varepsilon n\mu_k)^{1/k}$ for some large c, we can show that Theorem 1.6 simplifies to:

Theorem 1.8 (Theorem 3.9). *For any $\mathcal{P}$ with k-th central moment μ_k for some $k\ge 2$, and any $\alpha>0$, the ε-DP mechanism in Theorem 1.6 takes*

$$n=\tilde{O}\left(\frac{1}{\varepsilon}\log\frac{|\mu|+(\varepsilon\mu_k)^{1/k}}{\varphi(1/16)}+\frac{\sigma^2}{\alpha^2}+\frac{\mu_k^{1/(k-1)}}{\varepsilon\alpha^{k/(k-1)}}\right) \tag{4}$$

samples and returns a $\tilde{\mu}$ such that $|\tilde{\mu}-\mu|\le\alpha$.

As our universal estimator does not need to know k and μ_k, Theorem 1.8 actually holds for any (k,μ_k), and the bound should really be the infimum over all k. In particular, if $\mathcal{P}$ is Gaussian, for which $\mu_k\le\sigma^k(k-1)!!$ for all k, Theorem 1.8 essentially degenerates into Theorem 1.7 by setting k to a large constant. Anyhow, we would still state Theorem 1.8 for a single k for ease of comparison with prior work. Also note that, as k gets smaller, the privacy term $\frac{\mu_k^{1/(k-1)}}{\varepsilon\alpha^{k/(k-1)}}$ becomes more significant compared with the sampling error $\frac{\sigma^2}{\alpha^2}$. This is intuitive: As $\mathcal{P}$ more spreads out, the individual values in the sample become more important, hence a higher cost for privacy. For $k=2$, the privacy term would dominate the sampling error for all $\varepsilon\le 1$.

For heavy-tailed distributions, the previous ε-mechanism [44] requires A1/A2 (for A2, their assumption is that $\mu_k\le\bar{\mu}_k\le R^k$ for given $k,\bar{\mu}_k$). Their sample complexity is

$$n=\tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\bar{\mu}_k^{1/k}}+\frac{\sigma^2}{\alpha^2}+\frac{\bar{\mu}_k^{1/(k-1)}}{\varepsilon\alpha^{k/(k-1)}}\right) \tag{5}$$

The sampling error term $\frac{\sigma^2}{\alpha^2}$ in (5) is the same as the one in (4). For the privacy term (the last term) in (5) to match that in (4), they will need $\bar{\mu}_k$ to be a constant-factor approximation of μ_k, which is not known how to obtain in a DP fashion. In fact, if $\mu_{2k}=\infty$, there is no way to obtain such a $\bar{\mu}_k$ even in the non-private setting other than by assumption. Assuming such a $\bar{\mu}_k=O(\mu_k)$ is given, it remains to compare $O\left(\log\frac{|\mu|+(\varepsilon\mu_k)^{1/k}}{\varphi(1/16)}\right)$ and $O\left(\log\frac{R}{\bar{\mu}_k^{1/k}}\right)=O\left(\log\frac{R}{\mu_k^{1/k}}\right)$. Since $|\mu|\le R, \mu_k^{1/k}\le R$, the former is always better unless $\mathcal{P}$ is ill-behaved: $\log\frac{1}{\varphi(1/16)}=\omega\left(\log\frac{1}{\mu_k^{1/k}}\right)$, i.e., $\varphi(1/16)$ is more than polynomially smaller than $\mu_k^{1/k}$. [44] also prove that $\Omega\left(\frac{\bar{\mu}_k^{1/(k-1)}}{\varepsilon\alpha^{k/(k-1)}}\right)$ samples are necessary. Similar to the argument in the Gaussian case, this lower bound is worst-case. It does not imply that this many samples are needed for every $\mathcal{P}$, or that the $\mu_k\le\bar{\mu}_k$ assumption is needed *a priori*.

Arbitrary distributions. If $\mathcal{P}$ only has finite $\mu_2=\sigma^2$, this corresponds to the most difficult distributions. Note that in this case, the sample complexity of [44] becomes

$$n=\tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma_{\max}}+\frac{\sigma^2}{\alpha^2}+\frac{\sigma_{\max}^2}{\varepsilon\alpha^2}\right)=\tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma_{\max}}+\frac{\sigma_{\max}^2}{\varepsilon\alpha^2}\right) \tag{6}$$

For this problem, [17] proposed a different mean estimator under A1/A2 with the sample complexity

$$n=\tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma_{\min}}+\frac{\sigma^2}{\varepsilon^2\alpha^2}+\frac{\sigma^2}{\varepsilon\alpha^2}\log\frac{R}{\sigma_{\min}}\right).^{7} \tag{7}$$

These two results do not dominate each other. If the given $\sigma_{\max}$ is a constant-factor approximation of σ, then (6) is better than (7); otherwise, (6) can be arbitrarily worse than (7). Note that again there is no way to obtain a good $\sigma_{\max}$ other than by assumption for a $\mathcal{P}$ with $\mu_4=\infty$.

Meanwhile, our algorithm is better than both [44] and [17] except for ill-behaved $\mathcal{P}$. Setting $k=2$, (4) becomes

$$n=\tilde{O}\left(\frac{1}{\varepsilon}\log\frac{|\mu|+\sqrt{\varepsilon}\sigma}{\varphi(1/16)}+\frac{\sigma^2}{\varepsilon\alpha^2}\right). \tag{8}$$

We have already compared with [44] above for a general k. For the comparison with [17], in addition to achieving pure DP, we see that the second term in (8) is strictly better than the last two terms in (7). The first term in (8) is also better than that in (7) in most reasonable cases, unless $\mathcal{P}$ is ill-behaved ($\varphi(1/16)\ll\sigma$) or a very small R is given (which would make the mean estimation problem meaningless).

1.3 Statistical Variance Estimation

For variance estimation, we first use the standard technique of randomly pairing up the elements in D. For each pair (X,X'), compute $Z=(X-X')^2$, and let $H=\{Z_1,Z_2,\cdots,Z_{n/2}\}$ be the resulting Z's. Since $\mathsf{E}[Z]=2\sigma^2$, the problem boils down to estimating $\mathsf{E}[Z]$. As our mean estimator is universal, we can apply it directly without worrying about the distribution of Z. In fact, the algorithm is even simpler, since the range of Z is zero-centered thus easier to find. The following is our main result on universal variance estimation:

Theorem 1.9 (Theorem 5.2 in our full version paper [25], informal). *There exists an ε-DP mechanism such that for any $\mathcal{P}$, given $D\sim\mathcal{P}^n$, if*

$$n>\Omega\left(\frac{1}{\varepsilon}\log\log\frac{1}{\varphi(1/16)}+\frac{1}{\varepsilon}\log\log(\mathrm{IQR})\right),$$

then it returns a $\tilde{\sigma}^2$ such that

$$|\sigma^2-\tilde{\sigma}^2|$$
$$=O\Big(\min_{\xi\ge 5\cdot\gamma(\varepsilon n)^2+2\sigma^2}\left(\left|\mathsf{E}\left[Z>2\sigma^2+\xi\right]\right|+\frac{\xi}{\varepsilon n}\log\log\frac{\gamma(\varepsilon n)}{\varphi(1/16)}\right)+\sqrt{\frac{\mu_4}{n}}\Big).$$

Going through similar exercises, we obtain simplified results in terms of the sample complexity for specific distributions.

[7] The result in [17] is claimed under CDP, which leads to a result under pure-DP by changing a distribution of noise.

Gaussian distributions. For Gaussian distributions, we have $\mu_4 = O(\sigma^4)$, and the simplified result is:

Theorem 1.10 (Theorem 5.3 in our full version paper [25]). *For any Gaussian $\mathcal{P}$, and any $\alpha > 0$, the ε-DP mechanism from Theorem 1.9 takes*

$$n = \tilde{O}\left(\frac{1}{\varepsilon}\max\left\{\log\log\sigma, \log\log\frac{1}{\sigma}\right\} + \frac{\sigma^4}{\alpha^2} + \frac{\sigma^2}{\varepsilon\alpha}\right) \tag{9}$$

samples and returns a $\tilde{\sigma}^2$ such that $|\tilde{\sigma}^2 - \sigma^2| \le \alpha$.

The last two terms are the same as for Gaussian mean estimation (Theorem 1.7), except that σ is replaced by σ^2. The first term is more interesting, where we are able to reduce a log term to a log log. This is exactly due to the simplification mentioned above: finding the width of the range enclosing $\mathsf{E}[Z]$ is exponentially easier than finding its location. Meanwhile, since the error in $\tilde{\sigma}^2$ is relative to σ^2 itself (in contrast, the error in $\tilde{\mu}$ is relative to σ), we have to prepare for the case where σ is very small, hence the $\log\log\frac{1}{\sigma}$ term in (9).

There are two existing Gaussian variance estimators that do not dominate each other. [45] under A1/A2 achieve a sample complexity of

$$n = \tilde{O}\left(\frac{1}{\varepsilon}\log\frac{R}{\sigma_{\min}} + \frac{1}{\varepsilon}\log\log\frac{\sigma_{\max}}{\sigma_{\min}} + \frac{\sigma^4}{\alpha^2} + \frac{\sigma^4}{\varepsilon\alpha^2}\right), \tag{10}$$

while [11, 41] under A2 achieve sample complexity

$$n = \tilde{O}\left(\frac{1}{\varepsilon}\log\frac{\sigma_{\max}}{\sigma_{\min}} + \frac{\sigma^4}{\alpha^2} + \frac{\sigma^2}{\varepsilon\alpha}\right). \tag{11}$$

These two results are incomparable: (11) has an (almost) quadratically better privacy term (the last term) than (10), but its dependency on $\frac{\sigma_{\max}}{\sigma_{\min}}$ is exponentially worse. On the other hand, (9) is better than both, unless A2 already gives a tight range for σ. In fact, if we are also given $\sigma_{\min}$, we can scale the data by $\frac{1}{\sigma_{\min}}$, and (9) would further simplify to $n = \tilde{O}\left(\frac{1}{\varepsilon}\log\log\frac{\sigma}{\sigma_{\min}} + \frac{\sigma^4}{\alpha^2} + \frac{\sigma^2}{\varepsilon\alpha}\right)$, which is always better than both (10) and (11).

Heavy-tailed distributions. Theorem 1.9 can be simplified into the following bound in terms of the central moments:

Theorem 1.11 (Theorem 5.5 in our full version paper [25]). *For any $\mathcal{P}$, and any $\alpha > 0$, the ε-DP mechanism in Theorem 1.9 takes $n = \tilde{O}\left(\frac{\mu_4}{\alpha^2} + \inf_{k\ge 4}\frac{\mu_k^{2/(k-2)}}{\varepsilon\alpha^{k/(k-2)}}\right)$ samples and returns a $\tilde{\sigma}^2$ such that $|\tilde{\sigma}^2 - \sigma^2| \le \alpha$.*

This is the first private variance estimator for heavy-tailed distributions.

1.4 IQR Estimation

Our IQR estimator is very simple: Discretize $\mathbb{R}$ using an appropriate bucket size return $\tilde{X}_{3n/4} - \tilde{X}_{n/4}$ using Theorem 1.5. We show that it achieves the following sample complexity:

Theorem 1.12 (Theorem 6.2 in our full version paper [25]). *There exists an ε-DP mechanism such that for any $\mathcal{P}$ and any $\alpha > 0$, it takes*

$$n = \tilde{O}\Big(\frac{1}{\varepsilon}\log\frac{|\mu| + \sigma + \gamma(n)}{\varphi(1/16)} + \frac{1}{\varepsilon\alpha\cdot\theta(\alpha/4)}\log\frac{\gamma(n)}{\varphi(1/16)} + \frac{1}{(\alpha\cdot\theta(\alpha/4))^2} + \frac{\mathrm{IQR}}{\alpha}\Big) \tag{12}$$

samples and returns an $\widetilde{\mathrm{IQR}}$ such that $|\widetilde{\mathrm{IQR}} - \mathrm{IQR}| \le \alpha$.

Here, $\theta(\alpha)$ is the average value of $f(x)$ in an interval of width α near $F^{-1}(1/4)$ and $F^{-1}(3/4)$ (formal definition given in our full version paper [25]). The previous IQR estimator [30] only satisfies (ε, δ)-DP. Their sample complexity is[8]

$$n = \tilde{O}\left(\frac{1}{\left(\theta(2n^{-1/3})\right)^6} + \frac{1}{\mathrm{IQR}^3} + \frac{1}{\alpha^3} + \exp\left(\frac{\mathrm{IQR}}{\varepsilon\alpha}\right)\right). \tag{13}$$

To simplify the comparison between (12) and (13), we consider a well-behaved $\mathcal{P}$ where $\theta(\alpha) = \Omega(1/\mathrm{IQR})$ (e.g., for Gaussians, we have $\theta(\alpha) = \Theta(1/\mathrm{IQR}) = \Theta(1/\sigma)$ for all $\alpha \le \mathrm{IQR}$) and ignore the logarithmic terms. Then (12) simplifies to $\tilde{O}\left(\frac{\mathrm{IQR}}{\varepsilon\alpha} + \frac{\mathrm{IQR}^2}{\alpha^2}\right)$ while (13) becomes $\tilde{O}\left(\frac{1}{\mathrm{IQR}^3} + \mathrm{IQR}^6 + \frac{1}{\alpha^3} + \exp\left(\frac{\mathrm{IQR}}{\varepsilon\alpha}\right)\right)$. Note that their sampling error $\mathrm{IQR}^6 + \frac{1}{\alpha^2} \ge (\mathrm{IQR}^6)^{1/3}\cdot(\frac{1}{\alpha^3})^{2/3} = \frac{\mathrm{IQR}^2}{\alpha^2}$, while their privacy term $\exp\left(\frac{\mathrm{IQR}}{\varepsilon\alpha}\right)$ is exponentially worse than ours. In particular, we get the right convergence rate $\alpha \propto 1/(\varepsilon n)$ for the privacy noise, which agrees with that for μ and σ^2. On the other hand, their rate is $\alpha \propto 1/(\varepsilon\log n)$.

1.5 Open Problems

The first open problem, obviously, is to extend our result to high dimensions. The challenge here is to achieve the optimal dependency on d (see Appendix A for more details). Another interesting direction is that, since the utility guarantees of our estimators depend on the parameters of $\mathcal{P}$ to be estimated, we cannot output confidence intervals. One possible solution is to derive privatized upper bounds of these parameters, but it may be challenging to make these upper bounds as tight as possible.

1.6 Organization

The paper is organized as follows. The formal definitions of certain concepts introduced above are given in Appendix B, together with some building blocks for our algorithm. In Section 2, we present our estimators in the empirical setting. In Section 3, we describe our universal mean estimator; estimators for variance and IQR are deferred to our full version paper [25]. Additional discussion of related work is also given in Appendix A.

2 EMPIRICAL ESTIMATORS

In this section, we design ε-DP mechanisms for estimating $\mu(D)$ and X_τ, where D is taken from $\mathbb{Z}$. We will first obtain $\tilde{\mathcal{R}}(D)$, a privatized $\mathcal{R}(D)$, and then invoke INV and the clipped mean estimator. It turns out that the instance optimality ratio crucially depends on how well $\tilde{\mathcal{R}}(D)$ approximates $\mathcal{R}(D)$. The extension to the continuous domain is given in our full version paper [25].

[8] [30] defines $\theta(\cdot)$ as the minimum value of $f(x)$ in a small interval near $F^{-1}(1/4)$ and $F^{-1}(3/4)$, but their proof still works even if it is defined as the average value, which makes the result stronger.

2.1 Estimate Radius

Before estimating $\mathcal{R}(D)$, we first estimate $\mathrm{rad}(D)$. We will show how to obtain a $\widetilde{\mathrm{rad}}(D)$ such that $\widetilde{\mathrm{rad}}(D) \le 2 \cdot \mathrm{rad}(D)$ while $[-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)]$ covers all but $O(\log\log(\mathrm{rad}(D)))$ elements of D.

Let $\mathrm{Count}(D, x) = |D \cap [-x, x]|$. It is easy to see that $\mathrm{Count}(\cdot, x)$ has the global sensitivity 1 for any x, while $\mathrm{rad}(D)$ is exactly the smallest x such that $\mathrm{Count}(D, x) \ge n$. Thus, a natural idea is to feed the query sequence $\mathrm{Count}(D, x)$ for $x = 0, 1, 2, 4, 8, \ldots$ to SVT with a threshold of $T = n$. However, doing so suffers from the "late stop" problem, i.e., SVT may stop at a $\widetilde{\mathrm{rad}}(D)$ that is too large due to the exponential growth rate of x. On the other hand, reducing the growth rate increases the length of the query sequence, degrading the utility of SVT. Inspired by Lemma B.9, we use $T = n - 6\log(2/\beta)/\varepsilon$ so that SVT will stop at the "right" place. The details are shown in Algorithm 8.

Algorithm 1: `InfiniteDomainRadius`.

Input: D, ε, β

$T = n - \frac{6}{\varepsilon}\log(2/\beta)$;
$\tilde{i} = \mathrm{SVT}\left(T, \varepsilon, \mathrm{Count}(D, 0), \mathrm{Count}(D, 2^0), \mathrm{Count}(D, 2^1), \ldots\right)$;
if $\tilde{i} = 1$ **then**
  $\widetilde{\mathrm{rad}}(D) = 0$;
else
  $\widetilde{\mathrm{rad}}(D) = 2^{\tilde{i}-2}$;
end
return $\widetilde{\mathrm{rad}}(D)$;

The privacy of `InfiniteDomainRadius` follows from that of the SVT and the post-processing property of DP. We analyze its utility below:

Theorem 2.1. *For any $D \in \mathbb{Z}^n$, with probability at least $1 - \beta$,* `InfiniteDomainRadius` *returns a $\widetilde{\mathrm{rad}}(D)$ such that $\widetilde{\mathrm{rad}}(D) \le 2 \cdot \mathrm{rad}(D)$ and*

$$\left|D \cap \overline{\left[-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)\right]}\right| = O\left(\frac{1}{\varepsilon}\log\left(\log\left(\mathrm{rad}(D)\right)/\beta\right)\right).$$

For the space limit, all proofs are moved to our full version paper [25].

2.2 Estimate Range

To find a good privatized range $\tilde{\mathcal{R}}(D)$, we first search for an $\tilde{X}$ that is very likely located inside $\mathcal{R}(D)$, which can be done using INV to find a privatized median over a finite domain, as most data have been covered in $[-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)]$. Next, we shift D to be centered around $\tilde{X}$, and run `InfiniteDomainRadius` again. The detailed algorithm is shown in Algorithm 7.

The privacy of `InfiniteDomainRange` follows from basic composition. Its utility is summarized by the following theorem:

Theorem 2.2. *Given ε, β, for any $D \in \mathbb{Z}^n$, if*

$$n > \frac{c_1}{\varepsilon}\log\left(\mathrm{rad}(D)/\beta\right),$$

Algorithm 2: `InfiniteDomainRange`.

Input: D, ε, β

$\widetilde{\mathrm{rad}}(D) = \texttt{InfiniteDomainRadius}(D, \frac{\varepsilon}{8}, \frac{\beta}{3})$;
$D' = \mathrm{Clip}\left(D, [-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)]\right)$;
$\tilde{X} =$ `FiniteDomainQuantile`$\left(D', \frac{n}{2}, \mathbb{Z} \cap \left[-\widetilde{\mathrm{rad}}(D), \widetilde{\mathrm{rad}}(D)\right], \frac{\varepsilon}{8}, \frac{\beta}{3}\right)$;
$D'' = D - \tilde{X}$;
$\widetilde{\mathrm{rad}}(D'') = \texttt{InfiniteDomainRadius}(D'', \frac{3\varepsilon}{4}, \frac{\beta}{3})$;
$\tilde{\mathcal{R}}(D) = [\tilde{X} - \widetilde{\mathrm{rad}}(D''), \tilde{X} + \widetilde{\mathrm{rad}}(D'')]$;
return $\tilde{\mathcal{R}}(D)$;

Algorithm 3: `InfiniteDomainMean`.

Input: D, ε, β

$\tilde{\mathcal{R}}(D) = \texttt{InfiniteDomainRange}(D, \frac{4\varepsilon}{5}, \frac{\beta}{2})$;
$\tilde{\mu}(D) = \mathrm{ClippedMean}(D, \tilde{\mathcal{R}}(D)) + \mathrm{Lap}\left(5|\tilde{\mathcal{R}}(D)|/(\varepsilon n)\right)$;
return $\tilde{\mu}(D)$;

where c_1 is a universal constant, then with probability at least $1 - \beta$, `InfiniteDomainRange` *returns a range $\tilde{\mathcal{R}}(D)$ such that*

$$|\tilde{\mathcal{R}}(D)| \le 4 \cdot \gamma(D),$$

and

$$\left|D \cap \overline{\tilde{\mathcal{R}}(D)}\right| = O\left(\frac{1}{\varepsilon}\log\left(\log\left(\gamma(D)\right)/\beta\right)\right).$$

2.3 Mean Estimation

With a good $\tilde{\mathcal{R}}(D)$, we can now do mean estimation over an infinite domain. The algorithm is shown in Algorithm 3. Its privacy follows from basic composition, while its utility guarantee is as follows:

Theorem 2.3. *Given ε, β, for any $D \in \mathbb{Z}^n$, if*

$$n > \frac{c_1}{\varepsilon}\log\left(\mathrm{rad}(D)/\beta\right),$$

where c_1 is a universal constant, then with probability at least $1 - \beta$, `InfiniteDomainMean` *returns a $\tilde{\mu}(D)$ such that*

$$|\tilde{\mu}(D) - \mu(D)| = O\left(\frac{\gamma(D)}{\varepsilon n}\log\left(\log\left(\gamma(D)\right)/\beta\right)\right).$$

In Appendix B.2 and B.3, we give the definition for inward-neighborhood optimality and show for the empirical mean $\mu(D)$, its lower bound $\mathcal{L}_{\text{in-nbr}}(D) = \Omega(\gamma(D)/n)$ for every D. This means `InfiniteDomainMean` is inward-neighborhood optimal with an optimality ratio of $c = O(\log\log(\gamma(D))/\varepsilon)$ for constant β. Below, we show that this c is worst-case optimal in the finite-domain case. In particular, it implies that the optimality ratio cannot be independent of D.

Theorem 2.4. *For the empirical mean $\mu(D)$, given any ε, any integer $N \ge 1$, and any $n > \log\log_2(N)/\varepsilon$, for any ε-DP mechanism $\mathcal{M} : [N]^n \to \mathbb{R}$, there exists $D \in [N]^n$, such that*

$$\mathrm{Err}(\mathcal{M}, D) \ge \frac{\gamma(D)}{3\varepsilon n}\log\log_2(N).$$

Algorithm 4: `InfiniteDomainQuantile`.

Input: D, τ, ε, β

$\tilde{\mathcal{R}}(D) = \texttt{InfiniteDomainRange}(D, \frac{4\varepsilon}{5}, \frac{\beta}{2})$;

$D' = \text{Clip}\left(D, \tilde{\mathcal{R}}(D)\right)$;

$\tilde{X}_\tau = \texttt{FiniteDomainQuantile}\left(D', \tau, \tilde{\mathcal{R}}(D), \frac{\varepsilon}{5}, \frac{\beta}{2}\right)$;

return $\tilde{X}_\tau$;

2.4 Quantile Estimation

Similarly, to find a privatized quantile over an infinite domain, we invoke `FiniteDomainQuantile` with $\tilde{\mathcal{R}}(D)$. The algorithm is shown in Algorithm 4. Its privacy is straightforward, while achieving $O(\log(\text{rad}(D))/\varepsilon)$ rank error:

Theorem 2.5. *Given ε, β, for any $D \in \mathbb{Z}^n$ and any $1 \le \tau \le n$, if*

$$n > \frac{c_1}{\varepsilon} \log\left(\text{rad}(D)/\beta\right),$$

where c_1 is a universal constant, then with probability at least $1-\beta$, `InfiniteDomainQuantile` *returns a value $\tilde{X}_\tau$ such that*

$$X_{\tau-t} \le \tilde{X}_\tau \le X_{\tau+t},$$

where

$$t = O\left(\frac{1}{\varepsilon} \log\left(\gamma(D)/\beta\right)\right).$$

The rank error of `FiniteDomainQuantile` is instance-specific, and worst-case optimal in the finite-domain case, by a reduction from the *interior-point problem*. Here, given a dataset $D \in [N]^n$, we want to return any integer inside $\mathcal{R}(D)$. It has been shown that any ε-DP mechanism for the interior point problem requires $n = \Omega(\log(N)/\varepsilon)$ [9, 16]. Given a (finite-domain) quantile mechanism with rank error t, we would be able to solve the interior-point problem on datasets with $2t$ elements by returning the median. Thus $\Omega(\log(N)/\varepsilon)$ is also a lower bound on the rank error.

3 STATISTICAL MEAN ESTIMATION

In this section, we consider the statistical mean estimation problem, i.e., given an i.i.d. sample $D \sim \mathcal{P}^n$ for an arbitrary, unknown $\mathcal{P}$ over $\mathbb{R}$, we wish to estimate $\mu_{\mathcal{P}}$. The idea is conceptually simple: We first discrete $\mathbb{R}$ with an appropriate bucket size b; then we invoke the empirical mean estimator over $\mathbb{Z}$. For the first step, we find a lower bound on the IQR, denoted $\overline{\text{IQR}}$, as the bucket size. For the second step, it turns out that directly invoking the empirical mean estimator in Theorem 2.3 results in sub-optimal errors in the statistical setting; instead, we shall use a tighter range to do the clipping.

3.1 Estimate a Lower Bound for IQR

Prior work under A2 simply uses $b = \sigma_{\min}$ as the bucket size, which would be dominated by the sampling error. In the absence of $\sigma_{\min}$, we seek to obtain a privatized lower bound of IQR, since $\text{IQR} \le 4\sigma$. Furthermore, recall $\gamma\left(2, \frac{3}{4}\right) \le \text{IQR}$ (Appendix B.1), thus if we randomly draw two values X, X' from $\mathcal{P}$, then with probability at least $\frac{1}{4}$, we have

$$|X - X'| \le \text{IQR}.$$

Meanwhile, we do not want a bucket size too small. We thus relate $|X - X'|$ with $\varphi(\cdot)$.

Lemma 3.1. *For any $X, X' \in \mathcal{P}$, with probability at least $1 - \frac{1}{8}$, we have $\varphi\left(\frac{1}{16}\right) \le |X - X'|$.*

To amplify the success probability, we randomly group the elements in D into pairs (X, X') and let $G = \{Y_1, Y_2, \ldots, Y_{n'}\}$ where $n' = n/2$ and $Y_i = |X - X'|$ for each pair. Again, suppose $Y_1 \le \cdots \le Y_{n'}$. Then certain quantiles of G will satisfy our needs with probability $1 - \beta$. More precisely:

Lemma 3.2. *Given β, for any $D \in \mathcal{P}^n$, if $n > c_1 \log(1/\beta)$, where c_1 is a universal constant, then with probability at least $1 - \beta$, we have, $\varphi\left(\frac{1}{16}\right) \le Y_{\frac{5n'}{32}}$ and $Y_{\frac{7n'}{32}} \le \text{IQR}$.*

Therefore, we can find a quantile between $Y_{\frac{5n'}{32}}$ and $Y_{\frac{7n'}{32}}$, say $Y_{\frac{3n'}{16}}$, as $\overline{\text{IQR}}$. However, we cannot use `InfiniteDomainQuantile` here as we have not discretized $\mathbb{R}$ yet. To get out of this circular dependency, we obverse that we do not need a $\tilde{Y}_{\frac{3n'}{16}}$ with a small rank error; instead, a rough constant-factor approximation will do. Thus, the idea is to run two instances of SVT, one with increasing thresholds and one with decreasing thresholds, as detailed in Algorithm 10.

Algorithm 5: `EstimateIQRLowerBound`.

Input: D, ε, β

$n' = \frac{n}{2}$;

Construct G from D;

$\tilde{i} = \text{SVT}\left(\frac{3n'}{16}, \frac{\varepsilon}{2}, \text{Count}(G, 2^0), \text{Count}(G, 2^1), \text{Count}(G, 2^2), \ldots\right)$;

$\tilde{j} = \text{SVT}\left(-\frac{3n'}{16}, \frac{\varepsilon}{2}, -\text{Count}(G, 2^0), -\text{Count}(G, 2^{-1}), -\text{Count}(G, 2^{-2}), \ldots\right)$;

if $\tilde{i} > 1$ **then**

  $\overline{\text{IQR}} = 2^{\tilde{i}-2}$;

else

  $\overline{\text{IQR}} = 2^{-\tilde{j}}$;

end

return $\overline{\text{IQR}}$;

The privacy of `EstimateIQRLowerBound` is straightforward; we analyze its utility below:

Theorem 3.3. *Given ε, β, for any $D \sim \mathcal{P}^n$, if*

$$n > \frac{c_1}{\varepsilon} \log\log \frac{1}{\varphi(1/16)} + \frac{c_2}{\varepsilon} \log\log\left(\text{IQR}\right) + \frac{c_3}{\varepsilon} \log(1/\beta),$$

where c_1, c_2, c_3 are universal constants, then with probability at least $1 - \beta$, `EstimateIQRLowerBound` *returns an $\overline{\text{IQR}}$ such that,*

$$\frac{1}{4} \cdot \varphi\left(\frac{1}{16}\right) \le \overline{\text{IQR}} \le \text{IQR}.$$

3.2 General Algorithm and Error Analysis

We mentioned that directly invoking `InfiniteDomainMean` over D, even with a good bucket size, results sub-optimal errors in the statistical setting with respect to the dependency on ε. Here we give an intuitive explanation. Recall that in `InfiniteDomainMean`,

we find a privatized range $\tilde{\mathcal{R}}(D)$ and use it with the clipped mean estimator. The error comes from two sources: (1) There are $\tilde{O}(1/\varepsilon)$ clipped outliers, each contributing $\gamma(D)/n$ bias. (2) The Laplace noise is proportional to $|\tilde{\mathcal{R}}(D)|/(\varepsilon n) = O(\gamma(D)/(\varepsilon n))$. One should thus match the two parts of errors for an optimal overall error bound. In the empirical setting, as D is arbitrary, simply using $\gamma(D)/n$ as an upper bound on the bias from clipping each outlier is already the best one can do. In the statistical setting, however, since D is an i.i.d. sample, this upper bound is too pessimistic.

Therefore, in the statistical setting, we try to use a tighter $\tilde{\mathcal{R}}(D)$ to perform more aggressive clipping. The idea is to sub-sample m elements from D and obtain a privatized range on the sample D', denoted $\tilde{\mathcal{R}}(D')$. A smaller m corresponds to more aggressive clipping, which increases the bias but reduces the noise. The optimal choice of m will depend on $\mathcal{P}$, which is not possible for a universal estimator. Fortunately and somehow amazingly, $m = \varepsilon n$ turns out to be a choice that is good enough, and here is the intuition: By Theorem B.4, the privacy budget on finding $\tilde{\mathcal{R}}(D')$ can be amplified to $\varepsilon' \approx \varepsilon n/m$. Therefore, there are $\tilde{O}(1/\varepsilon') = \tilde{O}(m/(\varepsilon n))$ outliers in D' outside $\tilde{\mathcal{R}}(D')$. However, there is essentially no room for improvement when the number of outliers in D' is less than 1, i.e., it is sufficient to set $m \geq \varepsilon n$. When $m \geq \varepsilon n$, the number of outliers in D is roughly $\tilde{O}(m/(\varepsilon n)) \cdot n/m = \tilde{O}(1/\varepsilon n)$, which is fixed, while a smaller m reduces $|\tilde{\mathcal{R}}(D')|$.

Algorithm 6: `EstimateMean`.

Input: D, ε, β

$\overline{\text{IQR}} = \texttt{EstimateIQRLowerBound}(D, \frac{\varepsilon}{8}, \frac{\beta}{9})$;

Let D' be a sample of εn values from D;

$\varepsilon' = \log\left(\frac{e^{\varepsilon}-1}{\varepsilon} + 1\right)$;

$\tilde{\mathcal{R}}(D') = \texttt{InfiniteDomainRange}(D', \frac{3\varepsilon'}{4}, \frac{\beta}{9})$ with $b = \overline{\text{IQR}}$;

$\tilde{\mu} = \text{ClippedMean}\left(D, \tilde{\mathcal{R}}(D')\right) + \text{Lap}\left(8|\tilde{\mathcal{R}}(D')|/(\varepsilon n)\right)$;

return $\tilde{\mu}$;

With the intuition above, we present our statistical mean estimator, as shown in Algorithm 6. Its privacy follows from Theorem B.4 and basic composition. Before analyzing its error, we first state a standard result relating $\mathcal{P}$ with its truncated version:

Lemma 3.4. *Let $X \sim \mathcal{P}$ and $\xi \geq 0$, and let $\bar{X}$ be the following random variable:*

$$\bar{X} = \begin{cases} \mu - \xi, & \text{if } X < \mu - \xi; \\ X, & \text{if } \mu - \xi \leq X \leq \mu + \xi; \\ \mu + \xi, & \text{if } X > \mu + \xi. \end{cases}$$

Let $\bar{\mu}$ and $\bar{\sigma}^2$ denote the mean and variance of $\bar{X}$. Then,

$$\bar{\sigma} \leq \sigma,$$

and

$$\mu - \bar{\mu} = \mathsf{E}[X < \mu - \xi] + \mathsf{E}[X > \mu + \xi].$$

We are now ready to analyze the error of `EstimateMean`.

Theorem 3.5. *Given ε, β, for any $D \sim \mathcal{P}^n$, if*

$$n > \frac{c_1}{\varepsilon}\log\log\frac{1}{\varphi(1/16)} + \frac{c_2}{\varepsilon}\log\log(\text{IQR}) + \frac{c_3}{\varepsilon}\log\frac{1}{\beta} + \frac{c_4}{\varepsilon}\log\frac{|\mu| + \sigma + \gamma(\varepsilon n, \beta/9)}{\varphi(1/16)},$$

where c_1, c_2, c_3, and c_4 are universal constants, then with probability at least $1-\beta$, `EstimateMean` returns a value $\tilde{\mu}$ such that

$$|\mu - \tilde{\mu}| = O\Bigg(\min_{\xi \geq 10\cdot\gamma(\varepsilon n, \frac{\beta}{9}) + 2\sigma}\left(\left|\mathsf{E}\left[X < \mu - \xi\right] + \mathsf{E}\left[X > \mu + \xi\right]\right| + \frac{\xi}{\varepsilon n}\log\left(\frac{1}{\beta}\log\frac{\gamma(\varepsilon n, \beta/9)}{\varphi(1/16)}\right)\right) + \sigma\sqrt{\frac{\log(1/\beta)}{n}}\Bigg).$$

We first explain each term in the theorem before presenting its proof. The first two terms in the requirement of n are from finding the bucket size, and the last one is for estimating $\tilde{\mathcal{R}}(D')$. In the error bound, all the terms in the $\min_\xi$ are due to privacy, while the last term is the sampling error. We would like to emphasize that although the requirement on n and the error bound depend on $\mathcal{P}$ (they have to), the algorithm does not need any *a priori* assumptions on $\mathcal{P}$. Furthermore, some of the dependencies can be improved if certain assumptions are made on $\mathcal{P}$. For instance, if $\sigma_{\min}$ is given, then there is no need to find a bucket size and the first two terms in the requirement on n will disappear, while the $\varphi\left(\frac{1}{16}\right)$ in both the requirement on n and the error bound will be replaced by $\sigma_{\min}$.

3.3 Error Bounds for Specific Distribution Families

To facilitate the comparison with prior work, below we derive simplified (and possibly looser) versions of Theorem 3.5 for certain distribution families. These simplified bounds can be easily rewritten into the sample complexity results stated in Section 1. We also set β as $\frac{1}{3}$.

Gaussian distributions. For a Gaussian $\mathcal{P}$, we have $\varphi(\beta) = \Theta(\sigma)$, $\text{IQR} = \Theta(\sigma)$, and $\gamma(\varepsilon n, \beta/9) = O\left(\sigma\sqrt{\log(\varepsilon n)}\right)$ by the standard Gaussian tail bound. In addition, due to its symmetry, $\mathsf{E}[X < \mu - \xi] + \mathsf{E}[X > \mu + \xi] = 0$ for any ξ. Fixing $\xi = c\sigma\sqrt{\log(\varepsilon n)}$ for some large constant c, Theorem 3.5 simplifies into:

Theorem 3.6. *Given ε, β, for any $D \sim \mathcal{P}^n$, where $\mathcal{P}$ is a Gaussian distribution, if*

$$n > \frac{c_1}{\varepsilon}\log\log\sigma + \frac{c_2}{\varepsilon}\log\log\frac{1}{\sigma} + \frac{c_3}{\varepsilon}\log\frac{|\mu|}{\sigma},$$

where c_1, c_2, c_3 are universal constants, then

$$\text{Err}(\texttt{EstimateMean}, D) = O\left(\frac{\sigma}{\sqrt{n}} + \frac{\sigma}{\varepsilon n}\log\log\log(\varepsilon n)\sqrt{\log(\varepsilon n)}\right).$$

Heavy-tailed distributions. Now, we consider the case where $\mathcal{P}$ has a bounded k-th central moment μ_k. Note that $\sigma \leq \mu_k^{1/k}$. In addition, we can also bound $\gamma(m, \beta)$ in terms of μ_k:

Lemma 3.7. *For any m, β, and $k \geq 2$, $\gamma(m, \beta) \leq 2\left(\frac{m\mu_k}{\beta}\right)^{1/k}$.*

Plugging these bounds into Theorem 3.5 and setting $\xi = c \cdot (\varepsilon n \mu_k)^{1/k}$ for some large constant c, the requirement on n becomes

$$n > \frac{c_1}{\varepsilon}\log\log\frac{1}{\varphi(1/16)} + \frac{c_2}{\varepsilon}\log\log(\text{IQR}) + \frac{c_3}{\varepsilon}\log\frac{|\mu| + (\varepsilon\mu_k)^{1/k}}{\varphi(1/16)},$$

and the error bound changes to

$$\begin{aligned}&\mathrm{Err}(\texttt{EstimateMean}, D)\\ =&O\left(\frac{\sigma}{\sqrt{n}}+\frac{\mu_k^{1/k}}{(\varepsilon n)^{1-1/k}}\log\log\frac{(\varepsilon n\mu_k)^{1/k}}{\varphi(1/16)}\right.\\ &\left.+\left|\mathsf{E}\left[X<\mu-c\cdot(\varepsilon n\mu_k)^{1/k}\right]+\mathsf{E}\left[X>\mu+c\cdot(\varepsilon n\mu_k)^{1/k}\right]\right|\right).\end{aligned} \tag{14}$$

Now, we further analyze the last term in (14). We first derive a lemma similar to the one in [44]:

Lemma 3.8. *Let $\mathcal{P}$ be a distribution with a bounded μ_k. Given ξ and t such that $\xi \geq 2\,(\mu_k/t)^{1/(k-1)}$, we have*

$$|\mathsf{E}[X<\mu-\xi]+\mathsf{E}[X>\mu+\xi]|\leq t.$$

By setting $\xi = c\cdot(\varepsilon n\mu_k)^{1/k}$ for $c\geq 2$ and $t=\frac{\mu_k^{1/k}}{(\varepsilon n)^{1-1/k}}$, we have

$$\left|\mathsf{E}\left[X<\mu-c\cdot(\varepsilon n\mu_k)^{1/k}\right]+\mathsf{E}\left[X>\mu+c\cdot(\varepsilon n\mu_k)^{1/k}\right]\right|\leq\frac{\mu_k^{1/k}}{(\varepsilon n)^{1-1/k}}.$$

Plugging this bound into (14), we obtain:

Theorem 3.9. *Given ε, β, for any $D\sim\mathcal{P}^n$ and any k if*

$$n>\frac{c_1}{\varepsilon}\log\log\frac{1}{\varphi(1/16)}+\frac{c_2}{\varepsilon}\log\log(\mathrm{IQR})+\frac{c_3}{\varepsilon}\log\frac{|\mu|+(\varepsilon\mu_k)^{1/k}}{\varphi(1/16)},$$

where c_1, c_2, c_3 are universal constants, then

$$\mathrm{Err}(\texttt{EstimateMean}, D)=O\left(\frac{\sigma}{\sqrt{n}}+\frac{\mu_k^{1/k}}{(\varepsilon n)^{1-1/k}}\log\log\frac{(\varepsilon n\mu_k)^{1/k}}{\varphi(1/16)}\right).$$

ACKNOWLEDGMENTS

This work has been supported by HKRGC under grants 16201819, 16205420, and 16205422. We would also like to thank Yuchao Tao for some helpful initial discussions on the problem and the anonymous reviewers who have made valuable suggestions on improving the presentation of the paper.

REFERENCES

[1] Ishaq Aden-Ali, Hassan Ashtiani, and Gautam Kamath. 2021. On the sample complexity of privately learning unbounded high-dimensional gaussians. In *Algorithmic Learning Theory*. PMLR, 185–216.
[2] Kareem Amin, Travis Dick, Alex Kulesza, Andrés Munoz Medina, and Sergei Vassilvitskii. 2019. Differentially Private Covariance Estimation.. In *NeurIPS*. 14190–14199.
[3] Kareem Amin, Alex Kulesza, Andres Munoz, and Sergei Vassilvtiskii. 2019. Bounding user contributions: A bias-variance trade-off in differential privacy. In *International Conference on Machine Learning*. PMLR, 263–271.
[4] Galen Andrew, Om Thakkar, Brendan McMahan, and Swaroop Ramaswamy. 2021. Differentially private learning with adaptive clipping. *Advances in Neural Information Processing Systems* 34 (2021), 17455–17466.
[5] Myrto Arapinis, Diego Figueira, and Marco Gaboardi. 2016. Sensitivity of Counting Queries. In *International Colloquium on Automata, Languages, and Programming (ICALP)*.
[6] Hassan Ashtiani and Christopher Liaw. 2022. Private and polynomial time algorithms for learning gaussians and beyond. In *Conference on Learning Theory*. PMLR, 1075–1076.
[7] Hilal Asi and John C Duchi. 2020. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. *Advances in neural information processing systems* 33 (2020).
[8] Borja Balle, Gilles Barthe, and Marco Gaboardi. 2018. Privacy amplification by subsampling: Tight analyses via couplings and divergences. In *Advances in Neural Information Processing Systems*. 6277–6287.
[9] Amos Beimel, Shiva Prasad Kasiviswanathan, and Kobbi Nissim. 2010. Bounds on the sample complexity for private learning and private data release. In *Theory of Cryptography Conference*. Springer, 437–454.
[10] Amos Beimel, Kobbi Nissim, and Uri Stemmer. 2013. Private learning and sanitization: Pure vs. approximate differential privacy. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 363–378.
[11] Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. 2020. CoinPress: Practical Private Mean and Covariance Estimation. *Advances in Neural Information Processing Systems* 33 (2020).
[12] Gavin Brown, Marco Gaboardi, Adam Smith, Jonathan Ullman, and Lydia Zakynthinou. 2021. Covariance-aware private mean estimation without private covariance estimation. *Advances in Neural Information Processing Systems* 34 (2021).
[13] Mark Bun, Cynthia Dwork, Guy N Rothblum, and Thomas Steinke. 2018. Composable and versatile privacy via truncated cdp. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. 74–86.
[14] Mark Bun, Gautam Kamath, Thomas Steinke, and Steven Z Wu. 2019. Private hypothesis selection. *Advances in Neural Information Processing Systems* 32 (2019).
[15] Mark Bun, Kobbi Nissim, and Uri Stemme. 2016. Simultaneous private learning of multiple concept. In *Proc. Innovations in Theoretical Computer Science*.
[16] Mark Bun, Kobbi Nissim, Uri Stemmer, and Salil Vadhan. 2015. Differentially private release and learning of threshold functions. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*. IEEE, 634–649.
[17] Mark Bun and Thomas Steinke. 2019. Average-Case Averages: Private Algorithms for Smooth Sensitivity and Mean Estimation. In *Advances in Neural Information Processing Systems 32 (NeurIPS '19)*. Curran Associates, Inc., 181–191.
[18] T Tony Cai, Yichen Wang, and Linjun Zhang. 2021. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *The Annals of Statistics* 49, 5 (2021), 2825–2850.
[19] TH Hubert Chan, Kai-Min Chung, Bruce M Maggs, and Elaine Shi. 2019. Foundations of differentially oblivious algorithms. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 2448–2467.
[20] Kamalika Chaudhuri, Anand D Sarwate, and Kaushik Sinha. 2013. A Near-Optimal Algorithm for Differentially-Private Principal Components. *Journal of Machine Learning Research* 14 (2013).
[21] Anindya De. 2012. Lower bounds in differential privacy. In *Theory of cryptography conference*. Springer, 321–338.
[22] Wei Dong, Juanru Fang, Ke Yi, Yuchao Tao, and Ashwin Machanavajjhala. 2022. R2t: Instance-optimal truncation for differentially private query evaluation with foreign keys. In *Proceedings of the 2022 International Conference on Management of Data*. 759–772.
[23] Wei Dong, Yuting Liang, and Ke Yi. 2022. Differentially Private Covariance Revisited. *arXiv preprint arXiv:2205.14324* (2022).
[24] Wei Dong and Ke Yi. 2021. Residual Sensitivity for Deferentially Private Multi-Way Joins. In *Proc. ACM SIGMOD International Conference on Management of Data*.
[25] Wei Dong and Ke Yi. 2021. Universal Private Estimators. *arXiv preprint arXiv:2111.02598* (2021).
[26] Wei Dong and Ke Yi. 2022. A Nearly Instance-optimal Differentially Private Mechanism for Conjunctive Queries. In *Proceedings of the 41st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*. 213–225.
[27] John Duchi and Ryan Rogers. 2019. Lower bounds for locally private estimation via communication complexity. In *Conference on Learning Theory*. PMLR, 1161–1191.
[28] John C Duchi, Michael I Jordan, and Martin J Wainwright. 2013. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, 429–438.
[29] John C Duchi, Michael I Jordan, and Martin J Wainwright. 2018. Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.* 113, 521 (2018), 182–201.
[30] Cynthia Dwork and Jing Lei. 2009. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 371–380.
[31] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*. Springer, 265–284.
[32] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N Rothblum, and Salil Vadhan. 2009. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 381–390.
[33] Cynthia Dwork and Aaron Roth. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9, 3–4 (2014), 211–407.
[34] Cynthia Dwork, Kunal Talwar, Abhradeep Thakurta, and Li Zhang. 2014. Analyze gauss: optimal bounds for privacy-preserving principal component analysis. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. 11–20.

[35] Marco Gaboardi, Ryan Rogers, and Or Sheffet. 2019. Locally Private Mean Estimation: Z-test and Tight Confidence Intervals. In *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2545–2554.
[36] Moritz Hardt and Kunal Talwar. 2010. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing*. 705–714.
[37] Samuel B Hopkins, Gautam Kamath, and Mahbod Majid. 2022. Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 1406–1417.
[38] Ziyue Huang, Yuting Liang, and Ke Yi. 2021. Instance-optimal Mean Estimation Under Differential Privacy. *Advances in Neural Information Processing Systems* (2021).
[39] Noah Johnson, Joseph P Near, and Dawn Song. 2018. Towards practical differential privacy for SQL queries. *Proceedings of the VLDB Endowment* 11, 5 (2018), 526–539.
[40] Matthew Joseph, Janardhan Kulkarni, Jieming Mao, and Steven Z Wu. 2019. Locally Private Gaussian Estimation. *Advances in Neural Information Processing Systems* 32 (2019), 2984–2993.
[41] Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. 2019. Privately Learning High-Dimensional Distributions. In *Proceedings of the 32nd Annual Conference on Learning Theory (COLT '19)*. 1853–1902.
[42] Gautam Kamath, Argyris Mouzakis, Vikrant Singhal, Thomas Steinke, and Jonathan Ullman. 2022. A private and computationally-efficient estimator for unbounded gaussians. In *Conference on Learning Theory*. PMLR, 544–572.
[43] Gautam Kamath, Or Sheffet, Vikrant Singhal, and Jonathan Ullman. 2020. Differentially private algorithms for learning mixtures of separated gaussians. In *2020 Information Theory and Applications Workshop (ITA)*. IEEE, 1–62.
[44] Gautam Kamath, Vikrant Singhal, and Jonathan Ullman. 2020. Private mean estimation of heavy-tailed distributions. In *Conference on Learning Theory*. PMLR, 2204–2235.
[45] Vishesh Karwa and Salil Vadhan. 2018. Finite Sample Differentially Private Confidence Intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[46] Pravesh Kothari, Pasin Manurangsi, and Ameya Velingker. 2022. Private robust estimation by stabilizing convex relaxations. In *Conference on Learning Theory*. PMLR, 723–777.
[47] Ios Kotsogiannis, Yuchao Tao, Xi He, Maryam Fanaeepour, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau. 2019. PrivateSQL: a differentially private SQL query engine. *Proceedings of the VLDB Endowment* 12, 11 (2019), 1371–1384.
[48] Xiyang Liu, Weihao Kong, Sham Kakade, and Sewoong Oh. 2021. Robust and differentially private mean estimation. *Advances in Neural Information Processing Systems* 34 (2021).
[49] Xiyang Liu, Weihao Kong, and Sewoong Oh. 2022. Differential privacy and robust statistics in high dimensions. In *Conference on Learning Theory*. PMLR, 1167–1246.
[50] H Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. 2017. Learning differentially private recurrent language models. *arXiv preprint arXiv:1710.06963* (2017).
[51] Frank D McSherry. 2009. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*. 19–30.
[52] Arjun Narayan and Andreas Haeberlen. 2012. DJoin: Differentially private join queries over distributed databases. In *USENIX Symposium on Operating Systems Design and Implementation*. 149–162.
[53] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2007. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. 75–84.
[54] Catuscia Palamidessi and Marco Stronati. 2012. Differential Privacy for Relational Algebra: Improving the Sensitivity Bounds via Constraint Systems. In *QAPL*.
[55] Venkatadheeraj Pichapati, Ananda Theertha Suresh, Felix X Yu, Sashank J Reddi, and Sanjiv Kumar. 2019. AdaCliP: Adaptive clipping for private SGD. *arXiv preprint arXiv:1908.07643* (2019).
[56] Davide Proserpio, Sharon Goldberg, and Frank McSherry. 2014. Calibrating Data to Sensitivity in Private Data Analysis. *Proceedings of the VLDB Endowment* 7, 8 (2014).
[57] Or Sheffet. 2017. Differentially private ordinary least squares. In *International Conference on Machine Learning*. PMLR, 3105–3114.
[58] Adam Smith. 2011. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*. 813–822.
[59] Yuchao Tao, Xi He, Ashwin Machanavajjhala, and Sudeepa Roy. 2020. Computing Local Sensitivities of Counting Queries with Joins. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 479–494.
[60] Abhradeep Guha Thakurta and Adam Smith. 2013. Differentially private feature selection via stability arguments, and the robustness of the lasso. In *Conference on Learning Theory*. PMLR, 819–850.
[61] Jalaj Upadhyay. 2018. The Price of Privacy for Low-rank Factorization. In *NeurIPS*.
[62] Salil Vadhan. 2017. The complexity of differential privacy. In *Tutorials on the Foundations of Cryptography*. Springer, 347–450.

A OTHER RELATED WORK

Many works on mean estimators extend to higher dimensions, such as [1, 6, 11, 12, 14, 18, 37, 38, 41, 42, 44, 46, 49]. Using the idea of [38] but replacing Gaussian mechanism with Laplace mechanism, we can extend our pure-DP estimator to the multivariate case. However, it does not get the optimal privacy term $\tilde{O}(d/(\varepsilon n))$. In fact, the problem is open even under A1/A2/A3 (assuming multivariate Gaussians for A3). [44] achieve the optimal $\tilde{O}(d/(\varepsilon n))$ but their algorithm runs in exponential time; the mechanism in [37] runs in polynomial time but its privacy error is $\tilde{O}(\sqrt{d/(\varepsilon n)})$. Besides, [48] propose a solution for robust mean estimation under differential privacy. The mean estimation problem has also been studied in the *local model* of DP [27–29, 35, 40], which is also an interesting direction to look at.

Covariance estimation in high dimensions has also received a lot of attention. [11, 14, 41] consider multivariate Gaussian distributions and make similar boundedness assumptions like A1/A2. [1, 6, 42, 46, 49] do not need such assumptions but they relax the privacy notion to approximate DP. [2, 23] study the covariance for the data with bounded norms, which is even stronger than A1/A2. [20, 34, 57, 61] study private PCA or OLS, which can also be used to estimate covariance. However, they also assume that the data have bounded norms.

In the empirical setting, worst-case optimality does not make sense for functions whose global sensitivity is very large or ∞, which is the case for the empirical mean $\mu(D)$ where D is drawn from an unbounded domain. Instance-optimality is thus more suitable, but as pointed out by [7], strict instance-optimality is not possible, who therefore propose a natural relaxation by considering a small neighborhood. Nevertheless, for functions like $\mu(D)$, the neighborhood has to be restricted to avoid degeneration into worst-case optimality [38], as we explain in Appendix B.3. Besides, as mentioned in Section 1.1, our empirical estimator can be used to answer self-join-free aggregation queries in a relational database. Answering aggregation queries has also been extensively studied in database community [5, 22, 24, 26, 39, 47, 51, 52, 54, 56, 59]. For more details, please see [22].

B PRELIMINARIES

B.1 Notation

Given a multiset $D = \{X_1, \ldots, X_n\} \in \mathbb{R}^n$ (we reorder D such that $X_1 \le \cdots \le X_n$), we introduce the following notation: Its *support* is $\mathrm{supp}(D)$, *range* is $\mathcal{R}(D) = [X_1, X_n]$, *width* is $\gamma(D) = X_n - X_1$, and *radius* is $\mathrm{rad}(D) = \max_i |X_i|$. It is clear that $\mathcal{R}(D) \subseteq [-\mathrm{rad}(D), \mathrm{rad}(D)]$, hence $\gamma(D) \le 2 \cdot \mathrm{rad}(D)$, but $\mathrm{rad}(D)$ can be arbitrary larger than $\gamma(D)$. For any $\mathcal{S} \subseteq \mathbb{R}$, let

$$|D \cap \mathcal{S}| = |\{1 \le i \le n \mid X_i \in D \cap \mathcal{S}\}| .$$

Given a continuous probability distribution $\mathcal{P}$ over $\mathbb{R}$, in addition to μ, σ^2, IQR defined in Section 1, we also need the following quantities: For any $k \ge 2$, the *kth-central moment* is $\mu_k = \mathsf{E}_{X\sim P}[|X - \mu|^k]$. In particular, $\mu_2 = \sigma^2$. For any $\beta \in (0, 1)$, the width of the *highest*

density region at level β is

$$\varphi(\beta) = \inf\left\{a_2 - a_1 \,\middle|\, a_1, a_2 \in \mathbb{R}, a_2 > a_1, \int_{a_1}^{a_2} f(x)\,\mathrm{d}x \geq \beta\right\}.$$

We will only need $\varphi(\beta)$ for some constant β. Note that $\varphi(1/2) \leq \mathrm{IQR} \leq 4\sigma$ (the first inequality is by definition and the second is by Chebyshev's inequality). For most $\mathcal{P}$, the three quantities are close (e.g., for a Gaussian $\mathcal{P}$, the three are all within a constant factor from each other), although the gap can be arbitrarily large for an ill-behaved $\mathcal{P}$.

For any $m \in \mathbb{N}$ and $\beta \in (0,1)$, define the (m,β)-*statistical width* of $\mathcal{P}$ as

$$\gamma(m,\beta) = \inf\left\{\lambda \in \mathbb{R} \,\middle|\, \Pr_{D\sim\mathcal{P}^m}[\gamma(D) \geq \lambda] \leq \beta\right\}.$$

Note that

$$\gamma\left(2, \frac{3}{4}\right) \leq \mathrm{IQR} \leq \gamma\left(\log_{\frac{4}{3}}(2/\beta), \beta\right).$$

The first inequality is because for $X \sim \mathcal{P}$, with probability $\frac{1}{2}$, $X \in [F^{-1}(1/4), F^{-1}(3/4)]$; the second inequality follows from the fact that $X \in [-\infty, F^{-1}(1/4)]$ and $X \in [F^{-1}(3/4), \infty]$ each happens with probability $\frac{1}{4}$, plus a union bound.

For $X \in \mathcal{P}$ and any $x \in \mathbb{R}$, define

$$\mathsf{E}[X \lesseqgtr x] := \mathsf{E}_{X\sim\mathcal{P}}\left[(X-x)\mathbb{I}(X \lesseqgtr x)\right].$$

Finally, we introduce the following shorthand: For any $a, b \in \mathbb{R}$, let $[a \pm b] := [a-b, a+b]$. For interval $[l,r]$ and $b \in \mathbb{R}$, let $[l,r] \pm b := [l-b, r+b]$. Define $[N] := \{0, 1, \ldots, N\}$.

B.2 Differential Privacy

The DP definition has already been introduced in Section 1. The following two properties of DP are well-known:

Lemma B.1 (Post Processing [31]). *If $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ satisfies ε-DP and $\mathcal{M}' : \mathcal{Y} \to \mathcal{Z}$ is any randomized mechanism, then $\mathcal{M}'(\mathcal{M}(D))$ satisfies ε-DP.*

Lemma B.2 (Basic Composition [31]). *If $\mathcal{M}_1 : \mathcal{X}^n \to \mathcal{Y}$ satisfies ε_1-DP and $\mathcal{M}_2 : \mathcal{X}^n \times \mathcal{Y} \to \mathcal{Z}$ satisfies ε_2-DP, then $\mathcal{M}_2(D, \mathcal{M}_1(D))$ satisfies $(\varepsilon_1+\varepsilon_2)$-DP.*

For any function Q, its local sensitivity at D is

$$\mathrm{LS}_Q(D) = \sup_{D\sim D'} |Q(D) - Q(D')|$$

and the global sensitivity is

$$\mathrm{GS}_Q = \sup_D \mathrm{LS}_Q(D).$$

A basic pure DP mechanism is the Laplace mechanism:

Lemma B.3 (Laplace Mechanism). *The mechanism*

$$\mathcal{M}_Q(D) = Q(D) + \mathrm{Lap}(\mathrm{GS}_Q/\varepsilon)$$

preserves ε-DP, where $\mathrm{Lap}(\mathrm{GS}_Q/\varepsilon)$ is a random variable drawn from the Laplace distribution with scale $\mathrm{GS}_Q/\varepsilon$.

Below we omit the subscript Q if the context is clear.

We also need the following result, which shows that privacy can be amplified by sampling.

Theorem B.4 (Sampling Amplification [8]). *Let $\eta \in (0,1)$. Given an ε-DP mechanism $\mathcal{M}$, define $\mathcal{S}_\eta$ as the operation of sampling ηn samples from D without replacement, then $\mathcal{M}(\mathcal{S}_\eta(D))$ preserves $(\log(1+\eta(e^\varepsilon-1)))$-DP.*

Note that for small ε, $\log(1+\eta(e^\varepsilon-1)) \approx \eta\varepsilon$.

B.3 Optimality

The high-probability error of using $\mathcal{M}(D)$ to approximate $Q(D)$ is defined as

$$\mathrm{Err}(\mathcal{M}, D, \beta) = \inf\left\{\lambda \in \mathbb{R} \mid \Pr\left[|\mathcal{M}(D) - Q(D)| \leq \lambda\right] \geq 1-\beta\right\}.$$

We often take β as a constant, say $\beta = 1/3$; in this case we simply write $\mathrm{Err}(\mathcal{M}, D)$.

The Laplace mechanism is worst-case optimal. However, for any function Q with $\mathrm{GS} = \infty$, such as the empirical mean $\mu(D)$ when D is taken from an unbounded domain, this optimality notion is meaningless. For such a Q, *instance-optimality* is more appropriate and much stronger:

Definition B.5 (Instance-optimality). Define the per-instance lower bound:

$$\mathcal{L}_{\mathrm{ins}}(D) = \inf_{\mathcal{M}'} \mathrm{Err}(\mathcal{M}', D).$$

Then a DP mechanism $\mathcal{M}$ is c-instance-optimal if

$$\mathrm{Err}(\mathcal{M}, D) \leq c \cdot \mathcal{L}_{\mathrm{ins}}(D)$$

for every D, where c is the optimality ratio, which may depend on D.

Unfortunately, $\mathcal{L}_{\mathrm{ins}}(D) = 0$ for every D due to the trivial DP mechanism $\mathcal{M}'(\cdot) \equiv Q(D)$. Thus, instance-optimal DP mechanisms do not exist unless Q is trivial (i.e., $Q(D)$ is the same for all D). Thus, the following natural relaxation has been proposed:

Definition B.6 (Neighborhood-optimality [7, 26]). Define the neighborhood lower bound:

$$\mathcal{L}_{\mathrm{nbr}}(D) = \inf_{\mathcal{M}'} \sup_{D':D'\sim D} \mathrm{Err}(\mathcal{M}', D').$$

Then a DP mechanism $\mathcal{M}$ is c-neighborhood-optimal if

$$\mathrm{Err}(\mathcal{M}, D) \leq c \cdot \mathcal{L}_{\mathrm{nbr}}(D),$$

for every D.

[62] show that $\mathcal{L}_{\mathrm{nbr}}(D) = \Theta(\mathrm{LS}(D))$ for every D. For the empirical mean $\mu(D)$, we have $\mathrm{LS}(D) = \infty$, since one can change an element in D arbitrarily to obtain D'. Thus this relaxation is "too much". To fix the issue, the idea is to restrict the neighborhood:

Definition B.7 (Inward-neighborhood-optimality [38]). Define the inward-neighborhood lower bound:

$$\mathcal{L}_{\text{in-nbr}}(D) = \inf_{\mathcal{M}'} \max_{D':D\sim D', \mathrm{supp}(D')\subseteq\mathrm{supp}(D)} \mathrm{Err}(\mathcal{M}', D').$$

Then a DP mechanism $\mathcal{M}$ is c-inward-neighborhood-optimal if

$$\mathrm{Err}(\mathcal{M}, D) \leq c \cdot \mathcal{L}_{\text{in-nbr}}(D),$$

for every D.

Algorithm 7: SVT.

Input: T, ε, $Q_1(D), Q_2(D), \ldots$
$\tilde{T} \leftarrow T + \text{Lap}(2/\varepsilon)$;
for $i \leftarrow 1, 2, \ldots$ **do**
  $\tilde{Q}_i(D) \leftarrow Q_i(D) + \text{Lap}(4/\varepsilon)$;
  if $\tilde{Q}_i(D) > \tilde{T}$ **then**
    Break;
  end
end
return i;

Note that the restricted neighborhood is only concerned with the utility of $\mathcal{M}$, which still has to meet the standard privacy requirement over all $D \sim D'$.

For any function Q, $\mathcal{L}_{\text{in-nbr}}(D)$ is always finite, as D can only have a finite number of inward neighbors (thus $\sup_{D'}$ is replaced by $\max_{D'}$). In particular, for the empirical mean $\mu(D)$, we have $\mathcal{L}_{\text{in-nbr}}(D) = \Theta(\gamma(D)/n)$ [38].

B.4 The Sparse Vector Technique

The Sparse Vector Technique (SVT) [32] has as input a (possibly infinite) sequence of queries, $Q_1, Q_2, \ldots$, where each query has global sensitivity 1, and a threshold T. It aims to find the first query whose answer is above T. The detailed algorithm is given in Algorithm 8. The SVT has been shown to satisfy ε-DP and enjoy the following error guarantee, which says that it will not stop until it gets close to T.

Lemma B.8 ([33]). *Suppose there exists a k_1 less than the length of the query sequence such that for all $i = 1, \ldots, k_1$, $Q_i(D) \le T - \frac{8}{\varepsilon}\log(2k_1/\beta)$. Then with probability at least $1-\beta$,* SVT *returns an $i \ge k_1 + 1$.*

However, as will be clear later, we will actually need a complementary result that guarantees that SVT will stop in time. The following lemma gives such a result. More importantly, it also yields a utility guarantee on the returned query.

Lemma B.9. *If there exists a k_2 such that $Q_{k_2}(D) \ge T + \frac{6}{\varepsilon}\log(2/\beta)$, then with probability at least $1-\beta$,* SVT *returns an $i \le k_2$ such that $Q_i(D) \ge T - \frac{6}{\varepsilon}\log(2k_2/\beta)$.*

B.5 The Inverse Sensitivity Mechanism

The *inverse sensitivity mechanism* (INV) [7] answers a query Q with a discrete output range $\mathcal{Y}$. Given Q and D, it returns a $y \in \mathcal{Y}$ such that there exists D' not too far from D and $Q(D') = y$. Concretely, for any D and any $y \in \mathcal{Y}$, define the path length:

$$\text{len}(Q, D, y) = \min_{D'}\{d(D, D') : Q(D') = y\},$$

where $d(D, D')$ is the number of different elements between D and D'. INV instantiates the exponential mechanism with len as the score function:

$$\Pr(\text{INV}(Q, D) = y) = \frac{\exp\left(-\varepsilon \cdot \text{len}(Q, D, y)/2\right)}{\sum_{y' \in \mathcal{Y}} \exp\left(-\varepsilon \cdot \text{len}(Q, D, y')/2\right)}.$$

The utility of INV follows from that of the exponential mechanism:

Lemma B.10 ([7]). *For any D and β, with probability at least $1-\beta$,* INV *returns a y such that there exists a D' with $d(D, D') \le \frac{2}{\varepsilon}\log(|\mathcal{Y}|/\beta)$ and $Q(D') = y$.*

Algorithm 8: `FiniteDomainQuantile`.

Input: D, τ, $\mathcal{X}$, ε, β
if $\tau \le \frac{2}{\varepsilon}\log(|\mathcal{X}|/\beta)$ **then**
  $\tau' = \frac{2}{\varepsilon}\log(|\mathcal{X}|/\beta)$;
else if $\tau \ge n - \frac{2}{\varepsilon}\log(|\mathcal{X}|/\beta)$ **then**
  $\tau' = n - \frac{2}{\varepsilon}\log(|\mathcal{X}|/\beta)$;
else
  $\tau' = \tau$;
end
Run INV to find the τ'-quantile of D.

INV can be used to find a privatized quantile X_τ of D, if D are taken from a finite ordered domain $\mathcal{X}$, where $\text{len}(Q, D, y)$ is simply the number of elements of D that are between X_τ and y. Since $\text{len}(Q, D, y)$ only changes when y passes some element in D, the exponential mechanism can be implemented in $O(n)$ time (given D sorted) as opposed to $O(|\mathcal{Y}|)$. Some care has to be taken if τ is too close to 1 or n, in which case INV may return something arbitrarily bad. The details are shown in Algorithm 8, which enjoys a rank error guarantee:

Lemma B.11. *Given ε, β and a finite ordered domain $\mathcal{X}$, for any $D \in \mathcal{X}^n$ and any $1 \le \tau \le n$, if $n > \frac{4}{\varepsilon}\log(|\mathcal{X}|/\beta)$, then with probability at least $1-\beta$,* `FiniteDomainQuantile` *returns an $\tilde{X}_\tau$ such that*

$$X_{\tau - \frac{4}{\varepsilon}\log(|\mathcal{X}|/\beta)} \le \tilde{X}_\tau \le X_{\tau + \frac{4}{\varepsilon}\log(|\mathcal{X}|/\beta)}.$$

[7] also propose a continuous version of SVT and [58] uses a similar idea to estimate a quantile in a bounded real value domain. However, as the domain is infinite, those algorithms do not have any utility guarantee in the empirical setting.

B.6 The Clipped Mean Estimator

A standard idea for dealing with an unbounded domain is to clip all values into a bounded range $[l, r]$. Define

$$\text{Clip}(X, [l, r]) = \begin{cases} l, & \text{if } X < l; \\ X, & \text{if } l \le X \le r; \\ r, & \text{if } X > r. \end{cases}$$

Let

$$\text{Clip}(D, [l, r]) = \{\text{Clip}(X_i, [l, r]) \mid X_i \in D\}.$$

Then the clipped mean estimator is

$$\text{ClippedMean}(D, [l, r]) = \mu(\text{Clip}(D, [l, r])).$$

It is obvious that $\text{ClippedMean}(\cdot, [l, r])$ has global sensitivity $(r-l)/n$. Thus, $\text{ClippedMean}(D, [l, r]) + \text{Lap}\left(\frac{r-l}{\varepsilon n}\right)$ satisfies ε-DP.