Mean Estimation Under Heterogeneous Privacy: Some Privacy Can Be Free

Syomantak Chaudhuri and Thomas A. Courtade
University of California, Berkeley
{syomantak, courtade}@berkeley.edu

2023 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-7554-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISIT54713.2023.10206746

***Abstract*—Differential Privacy (DP) is a well-established framework to quantify privacy loss incurred by any algorithm. Traditional DP formulations impose a uniform privacy requirement for all users, which is often inconsistent with real-world scenarios in which users dictate their privacy preferences individually. This work considers the problem of mean estimation under heterogeneous DP constraints, where each user can impose their own distinct privacy level. The algorithm we propose is shown to be minimax optimal when there are two groups of users with distinct privacy levels. Our results elicit an interesting saturation phenomenon that occurs as one group's privacy level is relaxed, while the other group's privacy level remains constant. Namely, after a certain point, further relaxing the privacy requirement of the former group does not improve the performance of the minimax optimal mean estimator. Thus, the central server can offer a certain degree of privacy without any sacrifice in performance.**

I. Introduction

Privacy-preserving techniques in data mining and statistical analysis have a long history [1]–[3], and are increasingly mandated by laws such as the GDPR in Europe [4] and the California Consumer Privacy Act (CCPA) [5]. The current de-facto standard for privacy - Differential Privacy (DP) - was proposed by [6], [7]. Recent extensions of DP include Renyi-DP [8], Concentrated-DP [9], and Zero-Concentrated-DP [10].

Statistical problems like mean estimation under privacy constraints are important in real-world applications, and there is a need to understand the trade-off between accuracy and privacy. Most existing works consider a uniform privacy level for all users (see, e.g., [11]) and do not capture the heterogeneity in privacy requirements encountered in the real-world. Such heterogeneity frequently emerges as users balance their individual privacy options against the utility they desire from a service. Thus, a natural question arises: how should one deal with heterogeneous privacy for statistical tasks, such as optimal mean estimation? The effect of heterogeneity of privacy levels on accuracy is not well-understood; here, we make an effort to further the understanding of this trade-off by focusing on the mean estimation problem as a step in this direction.

We remind the readers that in the classical estimation problem without privacy constraints, the mean squared error decays as $1/n$, where n is the sample size. While the same decay is also observed under homogeneous DP constraints, the cost of DP is present in the second-order term, generally of form $1/(\epsilon n)^2$, where ϵ is the privacy level [12].

A. Our Contribution

We consider the problem of univariate mean estimation of bounded random variables under the Central-DP model with Heterogeneous Differential Privacy (HDP). While the proposed scheme for mean estimation can handle arbitrary heterogeneity in the privacy levels, we prove the minimax optimality of the algorithm for the case of two groups of users with a distinct privacy constraint for each group. This setting is particularly relevant for social media platforms, where studies have found two broad groups of users – one with high privacy sensitivity and another that does not care [13, Example 2]. This two-level setting is also a good first-order approximation to scenarios where users have some minimum privacy protections (e.g., ensured by legislation), but may also opt-in to greater privacy protections (e.g., the 'do not sell my information' option mandated by the CCPA). The setting also includes when one group's data is public, corresponding to some already known information. For the general case of every user having a distinct privacy level, experiments confirm the superior performance of our proposed algorithm over other methods. We view this work as a step in understanding the trade-off in the heterogeneity of privacy and accuracy; some directions for further investigation are outlined in Section V.

Out of a total of n users, a fraction f are in the first group, and the rest are in the second group. Every user in the first group has a privacy level of ϵ_1 and the second group has a privacy level of ϵ_2 ($\epsilon_2 \geq \epsilon_1$). As in homogeneous DP, one might expect better accuracy in mean estimation as ϵ_2 is increased keeping n, f, ϵ_1 fixed[1]. However, we show that after a certain critical value, increasing ϵ_2 provides no further improvement in the accuracy of our estimator. By matching upper and lower bounds, we show that this phenomenon is fundamental to the problem and not an artifact of our algorithm. As a corollary of this saturation phenomenon, having a public dataset ($\epsilon_2 \to \infty$) has no particular benefit for mean estimation. Thus, the central-server can advertise and offer extra privacy up to the critical value of ϵ_2 to the second group while not sacrificing the estimation performance.

We stress that our results do not assume the fraction f to be constant. For example, for a fixed n, one could take $f = 0$ or $f = 1$ to recover results known for the homogeneous DP setting. One could also consider f to depend on n; e.g., consider $1 - f = c/n$ and $\epsilon_2 \to \infty$ to denote a constant number

[1]In DP framework, higher ϵ corresponds to lower privacy.

c of public data samples as we increase the number of private samples. The authors are unaware of any previous result that considers this problem in such a level of generality over $n, f, \epsilon_1, \epsilon_2$. Further, many of the techniques in the literature for DP mean estimation obtain the $1/n^2$ and the $1/n$ terms separately in the lower bound [14], [15], which cannot give tight results in f that we show.

In Section II, we define the problem setting, state the main theorems, the proposed algorithm, along with an interpretation of the results. Experiments and other baseline methods are presented in Section IV to support the theoretical claims made in this work. Conclusions and possible future directions are outlined in Section V.

B. Related Work

Estimation error in the homogeneous DP case has been studied in great detail in recent years (see [12], [14], [16], [17]) under both the Central-DP model and the Local-DP (LDP) model. In the LDP model, users do not trust the central server and send their data through a noisy channel to the server to preserve privacy [18], [19]. Tasks like query release, estimation, learning, and optimization have been considered in the setting of a private dataset assisted by some public data [20]–[27]. Using a few public samples to estimate Gaussian distributions with unknown mean and covariance matrix is considered in [28]. The public samples eliminate the need for prior knowledge of the range of mean, but the effect on accuracy with more public samples is not considered. HDP for federated learning is considered in [29]. They remark that naively taking a linear combination of gradients in the proportion of the privacy levels is suboptimal and propose an SVD-based projected gradient algorithm. A general recipe for dealing with HDP is given by [30], but their idea of scaling the data using a shrinkage matrix induces a bias in the estimator. Further, their approach can not deal with public datasets.

Personalized Differential Privacy (PDP) is another term for HDP in literature. Reference [31] studied PDP and proposed a computationally expensive way to partition users into groups with similar privacy levels. For each partition, standard DP algorithms can be used with respect to the minimum ϵ in each group. It is not immediately clear how to use these partitions for tasks like mean estimation - if we consider taking a linear combination of the outputs of each partition, then what is the optimal linear combination? Further, for the special case of the two groups we consider, the partitions are already clearly the two groups themselves. An alternate method by [32] proposes a mechanism that samples users with high privacy requirements with less probability. While this is a general approach for dealing with heterogeneity, it is not optimal for mean estimation. Indeed, sub-sampling when $\epsilon_2 \to \infty$ corresponds to deleting the ϵ_1-private data. Reference [33] also consider HDP mean estimation under the assumption that the variance of the unknown distribution is known. However, as they mention, they add more noise than necessary for privacy since they are essentially performing LDP instead of the more powerful Central-DP technique. As a result, no saturation phenomenon can be deduced in their method due to the excessive noise added. Further, they do not provide a lower bound. The PDP setting for finite sets is consdiered in [32], [34] and they give algorithms inspired by the Exponential Mechanism [35]. References [36], [37] consider a heterogeneous privacy problem for recommendation systems. PDP in the LDP setting has been studied by [38] for learning the locations of users from a finite set of possible locations. A recent work by [39] considered a Bayesian setting with uniform privacy and heterogeneity in the number of samples and distribution of each user's data. Another line of work by [40], [41] consider a more general notion of DP which encompasses HDP. References [42]–[44] consider a hybrid model where some users are satisfied with the Central-DP model while other users prefer the LDP model.

Most closely related to the present work is [45], which considers the general HDP setting for mean estimation in the context of efficient auction mechanism design from a Bayesian perspective. While they encounter a saturation-type phenomenon in their algorithm, it cannot tightly characterize the saturation condition even in the case of two datasets with distinct privacy levels (see Section IV). They also assume that all the privacy levels are less than 1. This assumption is central to their upper and lower bounds; hence, one cannot draw conclusions when there is a public dataset. Section IV contains more comparisons of our proposed method with that of [45].

II. Problem Definition

We begin with some notation: non-negative real numbers will be denoted by $\mathbb{R}_{\geq 0}$. As we consider one-dimensional datapoints in our datasets, we use boldfaces, such as $\boldsymbol{x}$ to denote a dataset or, equivalently, a vector. Capital boldfaces, such as $\boldsymbol{X}$, denote a random dataset, i.e., a random vector. Vectors with subscript i, e.g. $\boldsymbol{x}_i$, refer to the i-th entry of the vector, while we use the notion $\boldsymbol{x}_i'$ for a vector differing from $\boldsymbol{x}$ at the i-th position.

A natural definition of heterogeneous Differential Privacy can be given as follows. Similar definitions were also considered in [30], [45].

Definition 1 (Heterogeneous Differential Privacy)**.** *A randomized algorithm $M : \mathcal{X}^n \to \mathcal{Y}$ is said to be $\boldsymbol{\epsilon}$-DP for $\boldsymbol{\epsilon} \in \mathbb{R}^n_{\geq 0}$ if*

$$\mathbb{P}\{M(\boldsymbol{x}) \in S\} \leq e^{\epsilon_i} \mathbb{P}\{M(\boldsymbol{x}_i') \in S\} \quad \forall i \in [n], \tag{1}$$

for all measurable sets $S \subseteq \mathcal{Y}$, where $\boldsymbol{x}, \boldsymbol{x}_i' \in \mathcal{X}^n$ are any two 'neighboring' datasets that differ arbitrarily in only the i-th component. Note that the probability is taken over the randomized algorithm conditioned on the given datasets $\boldsymbol{x}, \boldsymbol{x}_i'$, i.e., it is a conditional probability.

For concreteness, we consider the case $\mathcal{X} = [-0.5, 0.5]$ and let $\mathcal{P}$ denote the set of all distributions with support on $\mathcal{X}$. The extension to intervals of general length is straightforward. Under this privacy setting, we investigate the problem of estimating the sample mean from the user's data, where each

user's data is sampled I.I.D. from a distribution $P \in \mathcal{P}$ over $\mathcal{X}$ with mean denoted by $\mu_P \in [-0.5, 0.5]$ from here on. Each 'datapoint' corresponds to a user's data in $\mathcal{X}$, i.e., user i has a datapoint $\boldsymbol{x}_i$ and the user has a privacy requirement of ϵ_i. Each user sends their data and their privacy level to the central server and the server needs to respect the users' privacy level (Central-DP model). Let the set of all $\boldsymbol{\epsilon}$-DP estimators from $\mathcal{X}^n$ to $\mathcal{Y} = [-0.5, 0.5]$ be denoted by $\mathcal{M}_{\boldsymbol{\epsilon}}$. We consider the error metric as Mean-Squared Error (MSE) and are interested in characterizing the minimax estimation error. For an algorithm $M(\cdot) \in \mathcal{M}_{\boldsymbol{\epsilon}}$, let $E(M)$ denote the worst-case error attained by it,

$$E(M) = \sup_{P \in \mathcal{P}} \mathbb{E}_{\boldsymbol{X} \sim P^n, M(\cdot)}[(M(\boldsymbol{X}) - \mu_P)^2].$$

Let $L(\boldsymbol{\epsilon})$ denote the minimax estimation error given by

$$L(\boldsymbol{\epsilon}) := \inf_{M \in \mathcal{M}_{\boldsymbol{\epsilon}}} E(M). \tag{2}$$

Henceforth, we restrict our attention to the case where, out of a population of n users, a fraction f has a known and equal privacy requirement of ϵ_1, and the rest of the population has a known and equal privacy requirement of ϵ_2 ($\epsilon_1 \leq \epsilon_2$ without loss of generality). Thus, for the described case of two groups of users, we write $L(\epsilon_1, \epsilon_2, n, f)$ for $L(\boldsymbol{\epsilon})$ defined in (2).

The notation $\gtrsim$ or $\lesssim$ denotes inequalities that hold up to a multiplicative universal constant (independent of $n, f, \epsilon_1, \epsilon_2$).

A. Main Results

We characterize $L(\epsilon_1, \epsilon_2, n, f)$ by giving an upper and lower bound, tight up to constant factors. For convenience, we define two problem-dependent quantities,

$$R := 1 + \frac{8}{\epsilon_1^2 n f} \; ; \quad r := \frac{\epsilon_2}{\epsilon_1}.$$

We also define the averages, $\bar{\epsilon} := f\epsilon_1 + (1-f)\epsilon_2$, and $\overline{\epsilon^2} := f\epsilon_1^2 + (1-f)\epsilon_2^2$ (note $\overline{\epsilon^2} \neq \bar{\epsilon}^2$, in general). We assume $n \geq 1$ throughout this work.

Theorem 1 (Upper Bound)**.** *There exists an $\boldsymbol{\epsilon}$-DP algorithm M which attains:*
(A) if $1 \leq r \leq R$:

$$\begin{aligned} E(M) &\leq \min\left\{ \frac{\overline{\epsilon^2}}{4n\bar{\epsilon}^2} + \frac{2}{(n\bar{\epsilon})^2}, \frac{1}{4} \right\} \\ &= \min\left\{ \frac{fR + (1-f)r^2}{4n[f + (1-f)r]^2}, \frac{1}{4} \right\} ; \end{aligned} \tag{3}$$

(B) if $R \leq r$:

$$\begin{aligned} E(M) &\leq \min\left\{ \frac{R}{4n[f + (1-f)R]}, \frac{1}{4} \right\} \\ &= \min\left\{ \frac{nf\epsilon_1^2 + 8}{4n[nf\epsilon_1^2 + 8(1-f)]}, \frac{1}{4} \right\} . \end{aligned} \tag{4}$$

The algorithm which achieves the upper bound in Theorem 1 is outlined in Algorithm 1 and we refer to this algorithm as Affine Differentially-Private Mean (ADPM) in the rest of this work. A proof sketch of Theorem 1 is presented in Section III-A. The weight $\boldsymbol{w}$ used by ADPM for the case of two groups of users is given in Table I. ADPM is inspired by the technique used by [45]. Note that while we prove the optimality of ADPM for the case of two groups of users, the algorithm also works for a general $\boldsymbol{\epsilon}$-DP requirement. Even in the general $\boldsymbol{\epsilon}$-DP case, ADPM empirically outperforms other existing algorithms (see Section IV).

Theorem 2 (Lower Bound)**.** *The minimax estimation error defined in* (2) *satisfies:*
(A) if $1 \leq r \leq R$:

$$L(\epsilon_1, \epsilon_2, n, f) \gtrsim \min\left\{ \frac{\overline{\epsilon^2}}{4n\bar{\epsilon}^2} + \frac{2}{(n\bar{\epsilon})^2}, \frac{1}{4} \right\} ;$$

(B) if $R \leq r$:

$$L(\epsilon_1, \epsilon_2, n, f) \gtrsim \min\left\{ \frac{R}{4n[f + (1-f)R]}, \frac{1}{4} \right\} .$$

A proof sketch of Theorem 2 is given in Section III-B. Theorem 1 and Theorem 2 together characterize the minimax estimation error $L(\epsilon_1, \epsilon_2, n, f)$ up to constant factors, demonstrating optimality of ADPM (modulo universal constant factors).

Theorems 1 and 2 together demonstrate a fundamental saturation phenomenon of practical importance that occurs when ϵ_2 is large. In particular, for $\epsilon_2 \geq R\epsilon_1$, the accuracy of any optimal algorithm does not further improve (modulo constant factors) if ϵ_2 is increased. In other words, the central server gains no improvement in the accuracy of mean estimation if the group with the lower privacy level keeps lowering their privacy level after a certain point. Thus, the central server might as well offer a privacy level of $R\epsilon_1$ to this group of users at no cost to the server. This starkly contrasts the homogeneous-DP case, where the central server gains accuracy as the privacy level for everyone is lowered.

Algorithm 1 Affine Differentially Private Mean (ADPM)

procedure ADPM($\boldsymbol{\epsilon}, \boldsymbol{x}$)

$$\text{Solve: } \boldsymbol{w}^* = \begin{cases} \text{argmin} & \frac{\|\boldsymbol{w}\|_2^2}{4} + 2\|\boldsymbol{w}/\boldsymbol{\epsilon}\|_\infty^2 \\ \text{subject to:} & \boldsymbol{w} \succcurlyeq \boldsymbol{0}, \;\; \sum_{i=1}^n w_i = 1 \end{cases}$$

▷ $\boldsymbol{w}/\boldsymbol{\epsilon}$ is element-wise division
if $\frac{\|\boldsymbol{w}^*\|_2^2}{4} + 2\|\boldsymbol{w}^*/\boldsymbol{\epsilon}\|_\infty^2 > \frac{1}{4}$ **then**
  return 0
else
  $N \sim \text{Laplace}(\|\boldsymbol{w}^*/\boldsymbol{\epsilon}\|_\infty)$
  return $\langle \boldsymbol{w}^*, \boldsymbol{x} \rangle + N$
end if
end procedure

Remark 1. *From Table I, it is interesting to note that if we keep other parameters constant and increase ϵ_2 from ϵ_1 to ∞, then initially, the optimal affine estimator assigns more weight to the ϵ_2-dataset. This can be intuitively understood by considering that this dataset needs less privacy so we*

TABLE I
OPTIMAL WEIGHTS OBTAINED BY ADPM: w_i REFERS TO THE WEIGHTS ASSIGNED TO USERS OF GROUP i.

Condition	Optimal w_1	Optimal w_2
$\epsilon_2 \le R\epsilon_1$	$\epsilon_1/n\bar{\epsilon}$	$\epsilon_2/n\bar{\epsilon}$
$\epsilon_2 \ge R\epsilon_1$	$1/n[f+(1-f)R]$	$R/n[f+(1-f)R]$

give a higher weight to it in the estimator. However, arbitrarily increasing ϵ_2 should not increase the weight for its corresponding dataset since this comes at the cost of higher variance due to effectively ignoring the ϵ_1 dataset. Indeed, when ϵ_2 crosses the threshold of $R\epsilon_1$, there is no further change in the weights. It can be understood as 'saturating' the weight after this point. Even if $\epsilon_2 \to \infty$, one would clip the weights and offer $R\epsilon_1$-DP privacy for this non-private dataset. In other words, this privacy comes for free!

B. Interpreting the Bounds on $L(\epsilon_1, \epsilon_2, n, f)$:

Only ϵ_2-private dataset: This case can be realized in three different ways: $f = 0$, or $\epsilon_1 = \epsilon_2$, or $\epsilon_1 = 0$[2]. The first case implies $\frac{\epsilon_2}{\epsilon_1} \le R \to \infty$ so (3) gives an error of order $O(\frac{1}{n} + \frac{1}{(n\epsilon_2)^2})$, the minimax bound known for a homogeneous ϵ_2-DP mean estimation with n datapoints. For the second case, the same result applies as $1 = \frac{\epsilon_2}{\epsilon_1} < R$. In the third case, $\epsilon_2 \le \epsilon_1 R$ so the error is of order $O(\frac{1}{n(1-f)} + \frac{1}{(n(1-f)\epsilon_2)^2})$ - again matching the minimax bound for homogeneous ϵ_2-DP mean estimation with $n(1-f)$ datapoints. There are $n(1-f)$ datapoints since the algorithm needs to be independent of the ϵ_1-private data for $\epsilon_1 = 0$.

An ϵ_1-private dataset and a public dataset: Letting $\epsilon_2 \to \infty$ implies a completely public dataset. Keeping ϵ_1 fixed, since $\frac{\epsilon_2}{\epsilon_1} \ge R$, (4) yields an error of $\frac{nf\epsilon_1^2+8}{4n[nf\epsilon_1^2+8(1-f)]}$. At first glance, it behaves roughly like $\frac{1}{4n}$, corresponding to having n public samples. However, this is misleading since we care about the sharp dependence on f, ϵ_1 and the factor of $\frac{nf\epsilon_1^2+8}{nf\epsilon_1^2+8(1-f)}$ accounts for this, as we demonstrate next.

Tight in f: consider $\epsilon_1 \to 0$, and we get an error bound of $\frac{1}{4n(1-f)}$, which corresponds to the known bound for having $n(1-f)$ public samples (and thus, the bound is sharp in f). As a side note, depending on how we take the limits for $(\epsilon_1, \epsilon_2) \to (0, \infty)$, we may end up in case (A) in Theorem 1 as well but the upper bound is identical for both cases for this limit.

Tight in ϵ_1: Taking $f = 1$ yields an error of $O(\frac{1}{n} + \frac{1}{(n\epsilon_1)^2})$, the minimax rate for n users under homogeneous ϵ_1-DP.

III. PROOF SKTECHES

Detailed proofs can be found in [46].

A. Upper Bound

The randomized affine estimator $\langle \boldsymbol{x}, \boldsymbol{w} \rangle + L(\eta)$, where $L(\eta)$ is zero-mean Laplace noise with parameter η, can be shown to be $(\boldsymbol{w}/\eta)$-DP (see [46, Lemma 1]). Choosing $w_i = \epsilon_i / \|\boldsymbol{\epsilon}\|_1$ and $\eta = 1/\|\boldsymbol{\epsilon}\|_1$ is a possible way to satisfy the constraints - this suboptimal estimator proportionally weighs the datapoints based on (lack of) privacy requirement. If we have one datapoint with huge ϵ_2 and all the other $(n-1)$ datapoints have small ϵ_1, then proportional weighting will essentially use only one sample to estimate the mean and this leads to higher variance. Instead, we find the optimal weights $\boldsymbol{w}$ by optimizing the worst-case MSE.

[2]For $\epsilon_1 \to 0$, the saturation condition should be interpreted as $\epsilon_2 \le \epsilon_1 + 8/nf\epsilon_1$

B. Lower Bound

We use Le Cam's method specialized to differential privacy for proving the lower bound, based on ideas from [15], [19], [47]. Our method is similar to [45] but it is more robust since it can handle arbitrarily large ϵ values, as is required for the case when we have a public dataset. Intuitively, DP restricts the variation in output probability with varying inputs which helps bound the total-variation norm term in Le Cam's method.

IV. EXPERIMENTS

A. Baseline Schemes

We consider some baseline techniques for comparison and comment on why they are not optimal in HDP.

Uniformly enforce ϵ_1-DP (UNI): This approach offers ϵ_1 privacy to all the datapoints and uses the minimax estimator, i.e., the sample mean added with Laplace noise, to get an error of $O(1/n + 1/(n\epsilon_1)^2)$. UNI can be arbitrarily worse than the ADPM (consider a single low ϵ_1 datapoint).

Sampling Mechanism (SM): Based on the work of [32], let $t = \|\boldsymbol{\epsilon}\|_\infty$, and sample i-th datapoint independently with probability $(e^{\epsilon_i} - 1)/(e^t - 1)$. Apply homogeneous t-DP minimax estimator on the sub-sampled dataset. [32] proved this mechanism is $\boldsymbol{\epsilon}$-DP. However, when one dataset is public, the SM algorithm disregards the ϵ_1-private dataset.

Local Differential Private Estimator (LDPE): Consider the algorithm that combines the ϵ_1-DP and ϵ_2-DP mean estimates from the two datasets in a linear fashion. That is, it adds Laplace noise $L(\frac{1}{\epsilon_1 nf})$ to the sample mean of the first dataset and independent Laplace noise $L(\frac{1}{\epsilon_2 n(1-f)})$ to the sample mean of the second dataset, followed by optimal linear combinations of these two aggregates to minimize the mean squared error if the variance of the unknown distribution is known (see [33] for details). We take the worst-case variance as a proxy in our problem setting. When ϵ_1 and ϵ_2 are nearly the same, LDPE is worse since it adds more noise than necessary - this is a known shortcoming of the Local-DP model. ADPM scales better to the general case of $\boldsymbol{\epsilon}$-DP but LDPE is a decent baseline to compare it with. Note that LDPE is optimal as well when $\epsilon_2 \to \infty$ (see [46, Remark 2]).

FME [45]: For brevity, we direct the readers to [45, Theorem 1] for details on the algorithm. We refer to this algorithm as FME in the rest of this work. One of the shortcomings of this method is it assumes $\|\boldsymbol{\epsilon}\|_\infty \le 1$ for its theoretical guarantees. For our experiments, we still use this algorithm as it is stated for $\|\boldsymbol{\epsilon}\|_\infty > 1$. Even when $\|\boldsymbol{\epsilon}\|_\infty \le 1$, FME may assign much smaller weights than what ADPM does to the less private dataset [46, Example 1]. This might be one of the

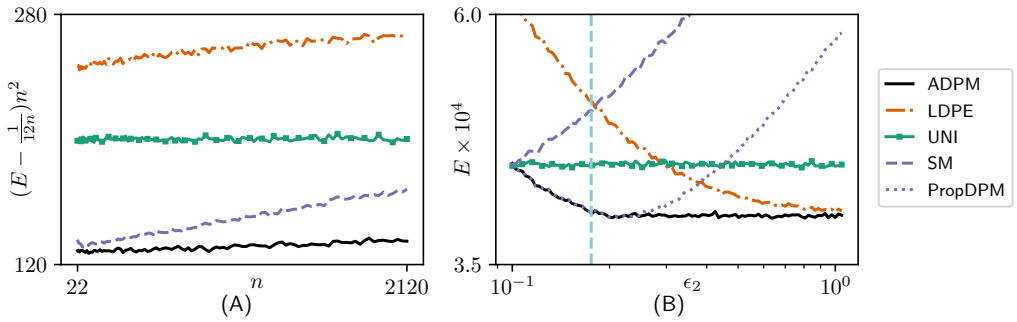

Fig. 1. We compare ADPM (our method) to other baseline methods in the above two graphs. FME is not plotted since its performance is an order worse than others in the above graphs; a comparison against FME is presented in [46]. Subfigure (A) plots $(E(M) - \frac{1}{12n})n^2$ vs n for each algorithm, keeping $\epsilon_1 = 0.1,\ \epsilon_2 = 0.15,\ f = 0.5$. Subfigure (B) plots $E(M) \times 10^4$ vs ϵ_2 while keeping $\epsilon_1 = 0.1,\ f = 0.7,\ n = 10^3$. The vertical dashed line marks the saturation point of ADPM at $\epsilon_2 = R\epsilon_1$. PropDPM shows the degradation in performance of the proportional weighting scheme for larger ϵ_2.

reasons why this method does not show much improvement when ϵ_2 is increased [46, Figure 3].

Proportional DP (PropDPM): We refer to the affine estimator with weights proportional to the $\boldsymbol{\epsilon}$ vector and appropriate Laplace noise as PropDPM, the shortcomings of this estimator is described in Section III-A.

B. Empirical Results

We compare ADPM (our method) to PropDPM, LDPE, FME, SM, and UNI. To prevent cluttering the graphs, we do not perform experiments for the stretching mechanism proposed by [30]. We can construct a case to demonstrate its sub-optimality since it is a biased estimator (see [46, Example 2]). Comparisons with FME are not presented here since FME is an order of magnitude worse than other algorithms; such experiments can be found in [46]. All simulations are averaged over 200K runs of the algorithms.

In Figure 1(A), we plot $(\text{MSE} - \frac{1}{12n})n^2$ vs n keeping $\epsilon_1, \epsilon_2, f$ constant for ADPM, LDPE, SM, and UNI. We plot $(\text{MSE} - \frac{1}{12n})n^2$ to show the second-order behavior of the considered algorithms. The simulations are run with the true underlying distribution being the uniform distribution on $\mathcal{X}$.

When n is small, R is large, and weights proportional to $\boldsymbol{\epsilon}$ are optimal. SM algorithm does something similar so it is close to ADPM in performance. However, since it sub-samples the data, its MSE decays slightly slower in the first order and we can see this by the upward trend in the graph. The fact that LDPE algorithm performs worse than ADPM or UNI is not surprising since for the case considered, ϵ_1 and ϵ_2 are quite close so the additional noise it adds contributes to its sub-optimality.

Another insightful experiment, presented in Figure 1(B), is to vary ϵ_2 while keeping other parameters fixed and comparing the MSE. The true underlying distribution for this experiment is the Bernoulli distribution on $\{-0.5, 0.5\}$. The curve for ADPM reinforces Theorem 1 as there is no improvement in the MSE upon increasing ϵ_2 above $R\epsilon_1$. Further, PropDPM performs worse after this critical point as we expected from the discussion in Section III-A. As $\epsilon_2 \to \infty$, LDPE would achieve the optimal error but it does not appear to have any saturation phenomenon. This can be attributed to the suboptimal way of adding noise inherent to Local-DP.

TABLE II
COMPARISON OF MSE FOR HIGH AND LOW VARIANCE IN $\boldsymbol{\epsilon}$.

Method	**log MSE** High Var($\boldsymbol{\epsilon}$)	**log MSE** Low Var($\boldsymbol{\epsilon}$)
ADPM	**-9.3**	**-8.1**
PropDPM	-9.0	-8.1
LDPE	-7.2	-1.3
SM	-6.5	-7.9
FME	-6.2	-6.2
UNI	-5.1	-7.1

Now consider the HDP setting in its full generality of arbitrary $\boldsymbol{\epsilon}$. The minimization in ADPM can be solved efficiently by modern solvers. We consider two cases for $\boldsymbol{\epsilon}$ of size 10^3 - high variance and low variance in $\boldsymbol{\epsilon}$. The low variance case was obtained by uniformly sampling $\log \boldsymbol{\epsilon}$ in $[-3, -2]$. Independently, the high variance case corresponds to sampling $\log \boldsymbol{\epsilon}$ in $[-4, 2]$. Keeping the sampled $\boldsymbol{\epsilon}$ fixed, the average of the squared errors was taken over 20K simulations under Beta$(2, 3)$ distribution on $\mathcal{X}$. The results are presented in Table II. Unsurprisingly, UNI, PropDPM and ADPM enjoy similar performance in the low variance regime, while diverging in the higher variance regime.

V. CONCLUSION

We study the problem of mean estimation of bounded random variables under Heterogeneous Differential Privacy and propose the ADPM algorithm. Under HDP, when there are two groups of users with distinct privacy levels, we prove the minimax optimality of the algorithm. Experimentally our algorithm outperforms other methods even in the general HDP setting with many distinct privacy levels.

A line of future work that we are currently working on is to prove the optimality of our algorithm in the general setting of arbitrary $\boldsymbol{\epsilon}$ vector. The problem of mean estimation is also interesting in the unbounded setting under suitable assumptions such as sub-Gaussianity. Extending HDP for the multivariate case is another exciting avenue to consider.

ACKNOWLEDGEMENTS

We thank Justin Singh Kang and Yigit Efe Erginbas for valuable discussions. This work was supported in part by NSF CCF-1750430 and CCF-2007669.

REFERENCES

[1] L. J. Hoffman, "Computers and privacy: A survey," *ACM Computing Surveys*, vol. 1, no. 2, p. 85–103, June 1969.

[2] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 439–450.

[3] P. Ram Mohan Rao, S. Murali Krishna, and A. Siva Kumar, "Privacy preservation techniques in big data analytics: a survey," *Journal of Big Data*, vol. 5, pp. 1–12, 2018.

[4] EU, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (General Data Protection Regulation)," pp. 1–88, May 2016.

[5] CA, "California Consumer Privacy Act (CCPA)," *Office of the Attorney General, California Department of Justice*, 2018.

[6] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Annual international conference on the theory and applications of cryptographic techniques*. Springer, 2006.

[7] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of cryptography conference*. Springer, 2006.

[8] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE, 2017.

[9] C. Dwork and G. N. Rothblum, "Concentrated differential privacy," *arXiv preprint arXiv:1603.01887*, 2016.

[10] M. Bun and T. Steinke, "Concentrated differential privacy: Simplifications, extensions, and lower bounds," in *Theory of Cryptography Conference*. Springer, 2016.

[11] T. Wang, X. Zhang, J. Feng, and X. Yang, "A comprehensive survey on local differential privacy toward data statistics and analysis," *Sensors*, 2020.

[12] T. T. Cai, Y. Wang, and L. Zhang, "The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy," *The Annals of Statistics*, 2021.

[13] I. Kotsogiannis, S. Doudalis, S. Haney, A. Machanavajjhala, and S. Mehrotra, "One-sided differential privacy," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 2020.

[14] G. Kamath, J. Li, V. Singhal, and J. Ullman, "Privately learning high-dimensional distributions," in *Conference on Learning Theory*. PMLR, 2019.

[15] R. F. Barber and J. C. Duchi, "Privacy and statistical risk: Formalisms and minimax bounds," 2014.

[16] S. B. Hopkins, G. Kamath, and M. Majid, "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*, 2022.

[17] G. Kamath and J. Ullman, "A primer on private statistics," *arXiv preprint arXiv:2005.00010*, 2020.

[18] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM Journal on Computing*, 2011.

[19] J. C. Duchi, M. J. Wainwright, and M. I. Jordan, "Minimax optimal procedures for locally private estimation," *Journal of the American Statistical Association*, 2016.

[20] R. Bassily, A. Cheu, S. Moran, A. Nikolov, J. Ullman, and S. Wu, "Private query release assisted by public data," in *International Conference on Machine Learning*. PMLR, 2020.

[21] R. Bassily, S. Moran, and A. Nandi, "Learning from mixtures of private and public populations," *Advances in Neural Information Processing Systems*, 2020.

[22] T. Liu, G. Vietri, T. Steinke, J. Ullman, and S. Wu, "Leveraging public data for practical private query release," in *International Conference on Machine Learning*. PMLR, 2021.

[23] N. Alon, R. Bassily, and S. Moran, "Limits of private learning with access to public data," *Advances in neural information processing systems*, 2019.

[24] A. Nandi and R. Bassily, "Privately answering classification queries in the agnostic pac model," in *Algorithmic Learning Theory*. PMLR, 2020.

[25] P. Kairouz, M. R. Diaz, K. Rush, and A. Thakurta, "(Nearly) dimension independent private erm with adagrad rates via publicly estimated subspaces," in *Conference on Learning Theory*. PMLR, 2021.

[26] E. Amid, A. Ganesh, R. Mathews, S. Ramaswamy, S. Song, T. Steinke, V. M. Suriyakumar, O. Thakkar, and A. Thakurta, "Public data-assisted mirror descent for private model training," in *International Conference on Machine Learning*. PMLR, 2022.

[27] D. Wang, H. Zhang, M. Gaboardi, and J. Xu, "Estimating smooth glm in non-interactive local differential privacy model with public unlabeled data," in *Algorithmic Learning Theory*. PMLR, 2021.

[28] A. Bie, G. Kamath, and V. Singhal, "Private estimation with public data," in *Advances in Neural Information Processing Systems*, 2022.

[29] J. Liu, J. Lou, L. Xiong, J. Liu, and X. Meng, "Projected federated averaging with heterogeneous differential privacy," *Proceedings of the VLDB Endowment*, 2021.

[30] M. Alaggan, S. Gambs, and A.-M. Kermarrec, "Heterogeneous differential privacy," *Journal of Privacy and Confidentiality*, 2017.

[31] H. Li, L. Xiong, Z. Ji, and X. Jiang, "Partitioning-based mechanisms under personalized differential privacy," in *Pacific-asia conference on knowledge discovery and data mining*. Springer, 2017.

[32] Z. Jorgensen, T. Yu, and G. Cormode, "Conservative or liberal? personalized differential privacy," in *2015 IEEE 31St international conference on data engineering*, 2015.

[33] C. Ferrando, J. Gillenwater, and A. Kulesza, "Combining public and private data," in *NeurIPS 2021 Workshop Privacy in Machine Learning*, 2021.

[34] B. Niu, Y. Chen, B. Wang, J. Cao, and F. Li, "Utility-aware exponential mechanism for personalized differential privacy," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020.

[35] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 2007.

[36] Y. Li, S. Liu, J. Wang, and M. Liu, "A local-clustering-based personalized differential privacy framework for user-based collaborative filtering," in *International Conference on Database Systems for Advanced Applications*. Springer, 2017.

[37] S. Zhang, L. Liu, Z. Chen, and H. Zhong, "Probabilistic matrix factorization with personalized differential privacy," *Knowledge-Based Systems*, 2019.

[38] R. Chen, H. Li, A. K. Qin, S. P. Kasiviswanathan, and H. Jin, "Private spatial data aggregation in the local setting," in *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, 2016.

[39] R. Cummings, V. Feldman, A. McMillan, and K. Talwar, "Mean estimation with user-level privacy under data heterogeneity," *Advances in Neural Information Processing Systems*, vol. 35, pp. 29 139–29 151, 2022.

[40] S. Torkamani, J. B. Ebrahimi, P. Sadeghi, R. G. D'Oliveira, and M. Médard, "Heterogeneous differential privacy via graphs," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022.

[41] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, "Broadening the scope of differential privacy using metrics," in *International Symposium on Privacy Enhancing Technologies Symposium*. Springer, 2013.

[42] B. Avent, A. Korolova, D. Zeber, T. Hovden, and B. Livshits, "BLENDER: Enabling local search with a hybrid differential privacy model," in *26th USENIX Security Symposium*, 2017.

[43] B. Avent, Y. Dubey, and A. Korolova, "The power of the hybrid model for mean estimation," *Proceedings on Privacy Enhancing Technologies*, vol. 2020, pp. 48–68, 10 2020.

[44] A. Beimel, A. Korolova, K. Nissim, O. Sheffet, and U. Stemmer, "The Power of Synergy in Differential Privacy: Combining a Small Curator with Local Randomizers," in *1st Conference on Information-Theoretic Cryptography (ITC 2020)*, 2020.

[45] A. Fallah, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Optimal and differentially private data acquisition: Central and local mechanisms," in *Proceedings of the 23rd ACM Conference on Economics and Computation*. Association for Computing Machinery, 2022.

[46] S. Chaudhuri and T. A. Courtade, "Mean estimation under heterogeneous privacy: Some privacy can be free," *arXiv preprint*, 2023.

[47] J. Duchi, M. Jordan, and M. Wainwright, "Local privacy and minimax bounds: Sharp rates for probability estimation," *Adv. Neural Inform. Process. Syst*, 2013.

[48] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.