

Brendan Avent, Yatharth Dubey, and Aleksandra Korolova

The Power of the Hybrid Model for Mean Estimation

Abstract: We explore the power of the hybrid model of differential privacy (DP), in which some users desire the guarantees of the local model of DP and others are content with receiving the trusted-curator model guarantees. In particular, we study the utility of hybrid model estimators that compute the mean of arbitrary real-valued distributions with bounded support. When the curator knows the distribution’s variance, we design a hybrid estimator that, for realistic datasets and parameter settings, achieves a constant factor improvement over natural baselines. We then analytically characterize how the estimator’s utility is parameterized by the problem setting and parameter choices. When the distribution’s variance is unknown, we design a heuristic hybrid estimator and analyze how it compares to the baselines. We find that it often performs better than the baselines, and sometimes almost as well as the known-variance estimator. We then answer the question of how our estimator’s utility is affected when users’ data are not drawn from the same distribution, but rather from distributions dependent on their trust model preference. Concretely, we examine the implications of the two groups’ distributions diverging and show that in some cases, our estimators maintain fairly high utility. We then demonstrate how our hybrid estimator can be incorporated as a sub-component in more complex, higher-dimensional applications. Finally, we propose a new privacy amplification notion for the hybrid model that emerges due to interaction between the groups, and derive corresponding amplification results for our hybrid estimators.

Keywords: differential privacy, hybrid model, mean estimation

DOI 10.2478/popets-2020-0062

Received 2020-02-29; revised 2020-06-15; accepted 2020-06-16.

Brendan Avent: University of Southern California, E-mail: bavent@usc.edu

Yatharth Dubey: Georgia Institute of Technology[†], E-mail: yatharthdubey7@gatech.edu

Aleksandra Korolova: University of Southern California, E-mail: korolova@usc.edu

[†] Work done while visiting University of Southern California.

1 Introduction

Differential privacy [25], has become one of the de facto standards of privacy in computer science literature, particularly for privacy-preserving statistical data analysis and machine learning. Two traditional models of trust in DP literature are: the trusted-curator model (TCM) and the local model (LM). In the TCM, the curator receives the users’ true data and applies a randomized perturbation to achieve DP. In the LM, the curator receives users’ privatized data through a locally randomizing oracle that individually ensures DP for each user.

When it comes to deployments of DP, curators (e.g., companies, social scientists, government organizations) and users alike find the LM to be a better match for their privacy goals [34, 44]. Users’ privacy is assured even when they don’t trust the curator, and the curator limits its liability in the face of data leaks. However, it is well understood theoretically and empirically that utility outcomes are far worse in the LM than in the TCM [11, 16, 22, 27, 38]. This poses a challenge for curators with smaller user bases than the tech giants – on the one hand, they want to guarantee local DP to their users; on the other hand, they won’t be able to gain much utility from the data if they do.

Until recently, these trust models were considered mutually exclusively. Recent work of Avent et al. [3] observed that it can be beneficial to consider a *hybrid model* in which the majority of the users desire privacy in the LM, but a small fraction of users are willing to contribute their data with TCM guarantees. Indeed, it is common in industry to have a small group of “early adopters” or “opt-in users” who are willing to trust the organization more than the average user [45]. The work of [3] demonstrated experimentally that in the hybrid model, one can develop algorithms that take advantage of the opt-in user data to improve utility for the task of local search. However, their results left open the questions of how much improvement can be gained compared with the LM, the dependence of improvement on the parameters (e.g., sample size, number of opt-in users, privacy level, etc.), and whether hybrid algorithms exist that improve over both the TCM and LM algorithms simultaneously for all parameters (as their proposed algorithm, BLENDER, was only able to achieve simultaneous improvement for *some* parameters). These are

precisely the questions we address in this work for the problem of mean estimation for bounded real-valued distributions – a well-studied problem in statistical literature due to its prevalence as a fundamental building block in solutions to more complex tasks.

Contributions: Our contributions are as follows.

- We initiate the study of mean estimation in the hybrid model, where users with bounded real-valued data self-partition into two groups based on their preferred trust model. We rigorously formalize this problem in a statistical framework (Section 2), making minimal distributional assumptions for user data and even allowing the groups to come from separate distributions.
- We define a family of hybrid estimators that utilize a generic class of DP mechanisms (Section 3). To evaluate the hybrid estimators’ relative quality, we detail two non-hybrid baseline estimators and theoretically analyze their relationship.
- When the groups have the same distribution and the curator knows its variance, we derive a hybrid estimator from this family and analytically quantify its utility (Section 4). First, we prove that it always outperforms both non-hybrid baselines. Second, we prove that for practical parameters, it outperforms both baselines by a factor of no greater than 2.286. Additionally, we empirically evaluate our hybrid estimator on realistic distributions, showing that it achieves high utility in practice.
- When the groups have the same distribution but the curator *doesn’t* know its variance, we derive another hybrid estimator from this family and analytically quantify the estimator’s utility (Section 5). We prove that it always outperforms at least one non-hybrid baseline, and we precisely determine the conditions under which it outperforms both. We empirically evaluate it on realistic distributions and find that it not only achieves high utility in practice, but is sometimes utility competitive with the known-variance case.
- Since users’ self-partitioning may induce a bias between the groups, we evaluate our analytic utility expressions in the cases where the groups’ distributions diverge (Section 6). We find that the hybrid estimator is robust to divergences in the variances of the groups’ distributions, but sensitive to divergences in the means of the groups’ distributions.
- To demonstrate how more complex algorithms can use our estimator as a sub-component, we design a hybrid K -means algorithm which uses the hybrid es-

timator to merge the intermediate results of two non-hybrid K -means algorithms (Section 7). We experimentally show that this algorithm is able to achieve utility on-par with the better of its two non-hybrid building blocks, even though its underlying hybrid estimator is not explicitly designed for this problem.

- We introduce a new privacy amplification notion for the hybrid model that stems from interaction between the groups (Section 8). We derive the amplification level that our hybrid estimator achieves, and show that this amplification is significant in practice.

2 Preliminaries

In this section, we present the requisite background on differential privacy, define the mean estimation problem setting, and then review related work.

2.1 Differential Privacy Background

In this background, we precisely define differential privacy, then describe two of the most popular DP mechanisms, and conclude with a discussion of trust models.

Formally, a mechanism \mathcal{M} is (ϵ, δ) -DP [25] if and only if for all neighboring databases D and D' differing in precisely one user’s data, the following inequality is satisfied for all possible sets of outputs $Y \subseteq \text{Range}(\mathcal{M})$:

$$\Pr[\mathcal{M}(D) \in Y] \leq e^\epsilon \Pr[\mathcal{M}(D') \in Y] + \delta.$$

A mechanism that satisfies $(\epsilon, 0)$ -DP is said to be ϵ -DP.

Two of the most popular DP mechanisms are the Laplace mechanism [25] and the Gaussian mechanism [24]. These mechanisms ensure DP for any dataset D evaluated under a real-valued function f by computing $\tilde{f}(D) = f(D) + Y$. For the Laplace mechanism, Y is a random variable drawn from the Laplace distribution with scale parameter $b = \Delta_1 f / \epsilon$ (yielding standard deviation $s = \sqrt{2}b$), and $\Delta_1 f = \max \|f(D) - f(D')\|_1$ over all neighboring D, D' . For the Gaussian mechanism, Y is drawn from the Gaussian distribution with standard deviation $s = \sqrt{2 \ln(1.25/\delta)} \Delta_2 f / \epsilon$, and $\Delta_2 f = \max \|f(D) - f(D')\|_2$ over all neighboring D, D' .

As discussed in the introduction, there are two classic trust models in DP, distinguished by their timing of when the privacy perturbation is applied. In the LM, user data undergoes a privacy-preserving perturbation before it is sent to the curator; in the TCM the curator

first collects all the data, and then applies a privacy-preserving perturbation. The hybrid model, first proposed in [3], enables algorithms to utilize a combination of trust models. Specifically, the hybrid model allows users to individually select between the TCM and LM based on their personal trust preferences.

2.2 Problem Setting

Statistical literature on mean estimation spans a wide range of assumptions and utility objectives, so we begin by stating ours.

There are n users, with each user $i \in [n]$ holding data x_i to be used in a differentially private computation. Users self-partition into the TCM or the LM group and, regardless of their group choice, are guaranteed the same level of DP. Thus, a user’s group choice only reflects their trust towards the curator. The fraction of users that opted-in to the TCM is denoted as $c \in (0, 1)$, while the remaining $(1 - c)$ fraction prefer the LM. We denote the two groups as indices in the sets $T = \{1, \dots, cn\}$ and $L = \{cn + 1, \dots, n\}$ respectively, such that $T \cup L = [n]$.

Users who opt-in to the TCM group (referred to as TCM users) have data x_i drawn iid from an unknown distribution \mathcal{D}_T with mean μ_T , variance σ_T^2 , and support on the subset of interval $[0, m_T]$. Users who chose the LM group (referred to as LM users) have data x_i drawn iid from an unknown distribution \mathcal{D}_L with mean μ_L , variance σ_L^2 , and support on the subset of interval $[0, m_L]$. Together, the groups’ distributions form a mixture distribution $\mathcal{D} = c\mathcal{D}_T + (1 - c)\mathcal{D}_L$ with mean $\mu = c\mu_T + (1 - c)\mu_L$, variance $\sigma^2 = c(\mu_T^2 + \sigma_T^2) + (1 - c)(\mu_L^2 + \sigma_L^2)$, and support on $[0, m]$ where $m = \max\{m_T, m_L\}$. Table 1 provides a summary of all notation introduced in this work.

We make minimal assumptions about these distributions, and the curator’s knowledge thereof, throughout the paper. Specifically, in Sections 4 and 5, we assume $\mathcal{D} = \mathcal{D}_T = \mathcal{D}_L$ and analyze the scenarios where the curator both does and doesn’t know \mathcal{D} ’s variance respectively. In Section 6, we lift this equal-distributions assumption and analyze the consequences of the groups’ distributions diverging.

Measuring Utility

Our goal is to design accurate estimators of the mean μ of the mixture distribution \mathcal{D} . To measure utility, we benchmark all estimators against the non-private empirical mean estimator.

Definition 2.1. *The non-private empirical mean estimator is:*

$$\hat{\mu} = \frac{1}{n} \sum_{i \in [n]} x_i = c\hat{\mu}_T + (1 - c)\hat{\mu}_L.$$

This choice of benchmark reflects the fact that we are interested in the *excess error* introduced by the privatization scheme, beyond the inherent error induced by a finite sample size. Concretely, we measure the absolute error of an estimator $\tilde{\mu}$ by explicitly computing the mean squared error between it and the empirical mean.

Definition 2.2. *The MSE between an estimator $\tilde{\mu}$ and the non-private empirical mean $\hat{\mu}$ is:*

$$\mathcal{E} = \text{MSE}(\tilde{\mu}, \hat{\mu}) = \mathbb{E}[(\tilde{\mu} - \hat{\mu})^2]$$

Since the non-private empirical benchmark is used to measure the MSEs of all estimators in this paper, we simply refer to it as the MSE of the estimator.

| Symbol | Usage |
|---|---|
| ϵ, δ | Differential privacy parameters |
| n | Total number of users |
| c | Fraction of users who opt-in to TCM |
| T, L | Set of users who opted-in to TCM and set of users who are using LM, respectively |
| \mathcal{D} | Mixture distribution of both groups’ data |
| μ, σ^2, m | Mean, variance, and maximum support of \mathcal{D} |
| \mathcal{D}_T | Distribution of TCM groups’ data |
| μ_T, σ_T^2, m_T | Mean, variance, and maximum support of \mathcal{D}_T |
| \mathcal{D}_L | Distribution of LM groups’ data |
| μ_L, σ_L^2, m_L | Mean, variance, and maximum support of \mathcal{D}_L |
| x_i | User i ’s private data drawn iid from its group’s distribution |
| $\tilde{\mu}, \hat{\mu}_T, \hat{\mu}_L$ | Empirical mean estimates with all users, with only the TCM users, and with only the LM users, respectively |
| \mathcal{E} | MSE of an estimator with respect to $\hat{\mu}$ |
| $\tilde{\mu}_T, \mathcal{E}_T$ | TCM-Only estimator and its MSE |
| $\tilde{\mu}_F, \mathcal{E}_F$ | Full-LM estimator and its MSE |
| $\tilde{\mu}_L, \mathcal{E}_L$ | LM-Only estimator and its MSE |
| $\tilde{\mu}_H(w), \mathcal{E}_H(w)$ | Hybrid estimator with weight w and its MSE |
| Y_T, s_T^2 | TCM-Only estimator’s privacy random variable and its variance |
| $Y_{L,i}, s_{L,i}^2$ | User i ’s local privacy random variable and its variance |
| n_{crit}, c_{crit} | n and c values that partition where $\mathcal{E}_T \leq \mathcal{E}_F$ |
| $R(\mathcal{E}), r(\mathcal{E})$ | Relative improvement of estimator with MSE \mathcal{E} over the best and worst non-hybrid baselines, respectively |

Table 1. Comprehensive list of notation.

2.3 Related Work

We first compare our paper to the closest related work in the hybrid model [3], then discuss other works on DP mean estimation in non-hybrid models, and conclude by discussing other work in hybrid trust models.

Comparison to BLENDER [3]

A shared goal of our work and [3] is to take advantage of the hybrid model; beyond that, our work is fundamentally different from theirs in several ways.

The works address different problems. Avent et al. studied the problem of local search, which is a specific problem instance of heavy-hitter identification and frequency estimation. BLENDER tackles the frequency estimation portion of the problem by estimating counts of boolean-valued data using a variant of randomized response [56]. Our work focuses on the conceptually simpler, but not strictly weaker, problem of mean estimation of real-valued data using a broad class of privatization mechanisms. Because of this, their methods aren't applicable in this work.

Both works compare against the same types of baselines in their respective problems, but reach very different conclusions. The baselines are: 1) using only the TCM group's data under the TCM, and 2) using all data under the LM. [3] experimentally evaluated BLENDER and found that it typically outperformed at least one of these baselines, and occasionally outperformed both. For our problem, we derive utility expressions which prove that not only does our estimator always outperform at least one of the baselines, but that under certain assumptions, it *always outperforms both*.

Since the hybrid model enables users to self-partition into groups based on their trust model preference, an important consideration for utility is whether the groups have the same data distribution. In BLENDER, it was assumed that they did. In this work, our setting allows for groups to have the same or different distributions, and we derive analytic results for both cases.

Finally, the works have different takes on the role of interaction between groups. BLENDER carefully utilizes inter-group interactivity to achieve high utility. In this work, our hybrid estimators have no inter-group interactivity; these estimators achieve high utility, demonstrating that such interactivity isn't always necessary for improving utility. Moreover, we find that our lack of interactivity can improve users' privacy guarantees

with respect to a specific type of adversary, whereas BLENDER's interactivity gives no such improvement.

Non-Hybrid Mean Estimation

In this work, we use simple non-hybrid baseline mean estimators to enable us to obtain exact finite-sample utility expressions. However, DP mean estimation of distributions under both the TCM and LM has been studied since the models' introductions [17, 21, 56], and continues to be actively studied to this day [2, 15, 18, 20, 22, 28, 29, 32, 37, 39–42]. The goal of mean estimation research under both models is to maximize utility while minimizing the sample complexity by making various distributional assumptions. Some assumptions are stronger than those made in this work, such as assuming the data is drawn from a narrow family of distributions. Other assumptions are weaker, such as requiring only that the mean lies within a certain range or that higher moments are bounded. Because of the complexity of the mechanisms and their reliance on the distributional assumptions in the related works, their utility expressions are typically bounds or asymptotic rather than exact. Since we need exact finite-sample utility expressions to precisely determine the utility of our hybrid estimator relative to the baselines, we are unable to use their estimators and assumptions. Nevertheless, the related works show a practically significant sample complexity gap between the TCM and LM in their respective settings, further motivating mean estimation in the hybrid model.

Other Works in Hybrid Trust Models

Several other works utilize a hybrid combination of trust models. Of these, the closest-related work is the concurrent work of Beimel et al. [12]. Their work examines precisely the same hybrid DP model as this work, the combined *trusted-curator/local* model, and has the same goal of understanding whether this hybrid model is more powerful than its composing models. To accomplish this goal, they perform mathematical analyses on several theoretical problems, deriving asymptotic bounds which show that it is possible to solve problems in the hybrid model which cannot be solved in the TCM or LM separately. Additionally, they show that there are problems which cannot be solved in the TCM or LM separately, and can be solved in the hybrid model, but only if the TCM and LM groups interact with each other. Finally, they analyze a problem which *does not* significantly benefit from the hybrid model: basic hy-

pothesis testing. They prove that if there exists a hybrid model mechanism that distinguishes between two distributions effectively, then there also exists a TCM or LM mechanism which does so nearly as effectively. This result implies a lack of power of the hybrid model for the problem of mean estimation in certain settings.

Beyond the trusted-curator/local hybrid model, there are multiple alternative hybrid models in DP literature. The most popular is the *public/private* hybrid model of Beimel et al. [13] and Zhanglong et al. [36]. In this model, most users desire the guarantees of differential privacy, but some users have made their data available for use without requiring any privacy guarantees. In this model, some works assume that DP is achieved in the TCM [35, 50], while others assume that DP is achieved in the LM [55, 58]. In both cases, the works show that by operating in the public/private hybrid model, one can significantly improve utility relative to either model separately. Recently, theoretical works [9, 10] have explored the limits of this model’s power via lower bounds on the sample complexity of fundamental statistical problems.

Another DP hybrid model recently introduced is the *shuffle* model, which was conceptually proposed by Bittau et al. [16] before being mathematically defined and analyzed for its DP guarantees by Cheu et al. [19] and Erlingsson et al. [26]. In this model, users privately submit their data under the LM via an anonymous channel to the curator. The anonymous channel randomly permutes the users’ contributions so that the curator has no knowledge of what data belongs to which user. This “shuffling” enables users to achieve improved DP guarantees over their LM guarantees in isolation. Several works have since improved the original analyses and expanded the shuffle model to achieve even greater improvements in the users’ DP guarantee [6, 7, 30–33].

3 DP Estimators

In this section, we introduce the baseline estimators in the classic DP models, describe how we compare new estimators against these baselines, and define the family of hybrid estimators that we will be working with.

3.1 Baseline Non-hybrid DP Estimators

To understand the utility of the hybrid model, we put it into context with the utility of non-hybrid approaches.

The most natural non-hybrid alternatives are: 1) only using the TCM group’s data under the TCM, and 2) using all the data under the LM. This is motivated directly by the decision that an analyst must make when choosing between these two models: 1) use only the data of the more-trusting users under the TCM so as to not violate the trust preferences of the remaining users, or 2) treat all users the same under the less-trusting LM.

For both baselines, we consider estimators which directly compute the empirical mean, then add 0-mean noise from an arbitrary distribution whose variance is calibrated to ensure DP under the respective model. For pure ϵ -DP, this typically corresponds to using the Laplace mechanism; for (ϵ, δ) -DP, this typically corresponds to using the Gaussian mechanism [24]. We derive all results for the generic noise-addition mechanisms, and we use the ϵ -DP Laplace mechanism for empirical evaluations.

TCM-Only Estimator

The stated consequence of using the TCM is that the LM group’s data cannot be used. Thus, we design an estimator for this model and refer to it as the “TCM-Only” estimator.

Definition 3.1. *The TCM-Only estimator is:*

$$\tilde{\mu}_T = \frac{1}{cn} \sum_{i \in T} x_i + Y_T,$$

where Y_T is a random variable with 0 mean and s_T^2 variance chosen such that DP is satisfied for all TCM users.

Lemma 3.2. *$\tilde{\mu}_T$ has expected squared error:*

$$\mathcal{E}_T = \frac{(1-c)^2}{cn} \sigma_T^2 + \frac{1-c}{n} \sigma_L^2 + s_T^2 + (\mu_T - \mu)^2.$$

Proof. See Appendix A.

This error has three components, $\frac{(1-c)^2}{cn} \sigma_T^2 + \frac{1-c}{n} \sigma_L^2$, s_T^2 , and $(\mu_T - \mu)^2$. The first component is the error induced by subsampling only the TCM users – we refer to this as the *excess sampling error*. The second component is the error due to DP – we refer to this as the *privacy error*. The third component is the *bias error* induced by the groups’ means differing.

Full-LM Estimator

Since the LM doesn’t require trust in the curator, the data of *all* users can be used under this model. We de-

sign an estimator for this model and refer to it as the “Full-LM” estimator.

Definition 3.3. *Suppose each user i privately reports their data as $x_i + Y_{L,i}$, where $Y_{L,i}$ is a random variable with 0 mean and s_L^2 variance chosen such that DP is satisfied for user i . The Full-LM estimator is then:*

$$\tilde{\mu}_F = \frac{1}{n} \sum_{i \in [n]} (x_i + Y_{L,i}),$$

Lemma 3.4. *$\tilde{\mu}_F$ has expected squared error:*

$$\mathcal{E}_F = \frac{s_L^2}{n}.$$

Proof. See Appendix A.

This error only consists of a single simple component: the privacy error. Since the entire dataset is used, there is no excess sampling error and no bias error.

3.2 Utility Over Both Baselines

While our absolute measure of an estimator’s utility is the MSE (discussed in Section 2.2), we are primarily interested in a hybrid estimator’s *relative* gain over the baseline estimators. Explicitly, given some hybrid estimator with MSE \mathcal{E} , we consider the following measure of relative improvement over the baseline estimators.

Definition 3.5. *The relative improvement of an estimator with MSE \mathcal{E} over the best baseline estimator is:*

$$R(\mathcal{E}) = \frac{\min\{\mathcal{E}_T, \mathcal{E}_F\}}{\mathcal{E}}.$$

This measure of relative improvement can be re-written to explicitly consider the regimes where each of the baseline estimators achieves the $\min\{\cdot\}$. That is, we determine the parameter configurations in which the TCM-only estimator is better/worse than the Full-LM estimator. Intuitively, we expect that when very few users opt-in to the TCM, the TCM-Only estimator’s large excess sampling error will overshadow its smaller privacy error (relative to the Full-LM estimator’s privacy error). This intuition is made precise by considering “critical values” of c and n that determine the regimes where each of the estimators yields better utility.

Lemma 3.6. *Let n_{crit} and c_{crit} be defined as follows.*

$$n_{crit} = \frac{cs_L^2 + (1-c)((1-c)\sigma_T^2 - c\sigma_L^2)}{c((\mu_T - \mu)^2 + s_T^2)} \quad (1)$$

$$c_{crit} = \begin{cases} \frac{\sigma_L^2}{\sigma_L^2 + s_L^2}, & \sigma_T = \sigma_L, \\ \frac{2\sigma_T^2 - \sigma_L^2 + s_L^2 - \sqrt{(\sigma_L^2 - s_L^2)^2 + 4s_L^2\sigma_T^2}}{2(\sigma_T^2 - \sigma_L^2)}, & \sigma_T \neq \sigma_L \end{cases} \quad (2)$$

We have that $\mathcal{E}_T \leq \mathcal{E}_F$ if and only if $c > c_{crit} \wedge n \leq n_{crit}$.

Proof. Directly reduce the system of inequalities constructed by $\mathcal{E}_T \leq \mathcal{E}_F$ in conjunction with the regions given by the valid parameter ranges. This immediately yields the result. \square

This characterization allows us to partition the definition of relative improvement into the behavior of each baseline estimator, re-written as follows.

Definition 3.7. *The relative improvement of an estimator with MSE \mathcal{E} over the best baseline estimator is:*

$$R(\mathcal{E}) = \frac{1}{\mathcal{E}} \cdot \begin{cases} \mathcal{E}_T & \text{if } c > c_{crit} \wedge n \leq n_{crit} \\ \mathcal{E}_L & \text{otherwise} \end{cases}$$

The behavior of these two cases further depends on the privacy mechanism used, as that dictates s_T and s_L . For example, when using the ϵ -DP Laplace mechanism in the homogeneous setting where both group means are μ and variances are σ^2 , these definitions of critical values and relative improvement become the following.

Lemma 3.8. *Adding ϵ -DP Laplace noise for privacy, define $c_{crit} = \frac{\epsilon^2\sigma^2}{2m^2 + \epsilon^2\sigma^2}$ and $n_{crit} = \frac{2m^2}{c(2cm^2 - (1-c)\epsilon^2\sigma^2)}$. We have that $\mathcal{E}_T \leq \mathcal{E}_F$ if and only if $c > c_{crit} \wedge n \geq n_{crit}$.*

Definition 3.9. *Adding ϵ -DP Laplace noise for privacy, the relative improvement of an estimator with MSE \mathcal{E} over the best baseline estimator is:*

$$R(\mathcal{E}) = \frac{1}{\mathcal{E}} \cdot \begin{cases} \frac{1-c}{cn}\sigma^2 + \frac{2m^2}{c^2n^2\epsilon^2} & \text{if } c > c_{crit} \wedge n \geq n_{crit} \\ \frac{2m^2}{n\epsilon^2} & \text{otherwise} \end{cases}$$

Thus, once the fraction of users opting-in to the TCM is large enough, the TCM-Only estimator has better MSE than the Full-LM estimator. In all other regimes, the Full-LM estimator has better MSE than the TCM-Only estimator. This matches the intuition.

Designing a hybrid estimator which outperforms *at least one* of these baselines in *all* regimes (i.e., for all settings of parameters μ, σ^2, n, c, m , etc.) is trivial, as is designing a hybrid estimator which outperforms *both* baselines

in *some* regimes. One challenge solved in this work is designing a hybrid estimator which provably outperforms *both* baselines across *all* regimes.

3.3 Convex Hybrid DP Estimator Family

Simply described, this family of hybrid estimators has the two groups independently compute their own private estimates of the mean, then directly combines them as a weighted average. The TCM group’s estimator is the TCM-Only estimator. The LM group’s estimator is almost the same as the Full-LM estimator, except now with reports only from the LM users. We refer to this as the “LM-Only” estimator, and briefly detour to define and analyze it.

Definition 3.10. *The LM-Only estimator is:*

$$\tilde{\mu}_L = \frac{1}{(1-c)n} \sum_{i \in L} (x_i + Y_{L,i}),$$

where, for each $i \in L$, $Y_{L,i}$ is a random variable with 0 mean and s_L^2 variance chosen such that DP is satisfied for user i .

Lemma 3.11. $\tilde{\mu}_L$, has expected squared error:

$$\mathcal{E}_L = \frac{c^2}{(1-c)n} \sigma_L^2 + \frac{c}{n} \sigma_T^2 + \frac{1}{(1-c)n} s_L^2 + (\mu_L - \mu)^2.$$

Proof. See Appendix A.

This estimator has excess sampling error, privacy error, and bias error. Since it has strictly greater error than the Full-LM estimator, it is not used as one of the baseline estimators.

We now define a family of convexly-weighted hybrid estimators parameterized by weight $w \in [0,1]$, which we will use throughout this paper. For any w , the hybrid estimator computes a convex combination of the independent TCM-Only and LM-Only estimators.

Definition 3.12. *The hybrid estimator, parameterized by $w \in [0,1]$, is:*

$$\tilde{\mu}_H(w) = w\tilde{\mu}_T + (1-w)\tilde{\mu}_L.$$

Lemma 3.13. $\tilde{\mu}_H(w)$ has expected squared error:

$$\begin{aligned} \mathcal{E}_H(w) &= \frac{(w-c)^2}{cn} \sigma_T^2 + \frac{(w-c)^2}{(1-c)n} \sigma_L^2 + w^2 s_T^2 + \frac{(1-w)^2}{(1-c)n} s_L^2 \\ &\quad + (w\mu_T + (1-w)\mu_L - \mu)^2. \end{aligned}$$

Proof. See Appendix A.

This estimator has all three types of error – excess sampling error, privacy error, and bias error – where the amounts of each error type depend on the weighting w .

4 Homogeneous, Known-Variance Setting

In this section, we design a hybrid estimator in the homogeneous setting which outperforms the baselines by carefully choosing a particular weighting for the hybrid estimator family from Definition 3.12. To choose such a weighting, we restrict our focus to the homogeneous setting, where both groups’ means are the same ($\mu = \mu_T = \mu_L$) and variances are the same ($\sigma^2 = \sigma_T^2 = \sigma_L^2$). Beyond simplifying the expressions we’re analyzing, the homogeneous setting eliminates bias error from our defined estimators, which removes any dependence on μ from the derived error expressions. This is important, since the curator’s goal is to learn μ from the data; thus, no particular knowledge of μ is assumed. Therefore, in the homogeneous setting, a weighting can be chosen by analyzing the hybrid estimator’s derived error expressions without needing any knowledge of μ . However, there is still excess sampling error for the estimators in this setting – in other words, error expressions still depend on the data variance σ^2 . Thus, in this section, we make the common assumption in statistical literature that σ^2 is known to the curator, and derive and analyze the optimal hybrid estimator from the convex family.

KVH Estimator

We now derive and analyze the “known-variance hybrid” (KVH) estimator by computing the optimal weighting w^* that minimizes $\mathcal{E}_H(w)$. This can be analytically computed and directly implemented by the curator, since each term of $\mathcal{E}_H(w)$ is known in this setting.

Definition 4.1. *The known-variance hybrid estimator in the homogeneous setting is:*

$$\tilde{\mu}_{KVH} = w^* \tilde{\mu}_T + (1-w^*) \tilde{\mu}_L,$$

where $w^* = \frac{c(\sigma^2 + s_L^2)}{\sigma^2 + c(ns_T^2(1-c) + s_L^2)}$ is obtained by minimizing $\mathcal{E}_H(w)$ with respect to w .

Lemma 4.2. $\tilde{\mu}_{KVH}$ has expected squared error:

$$\mathcal{E}_{KVH} = \frac{(w^* - c)^2}{c(1-c)n} \sigma^2 + w^{*2} s_T^2 + (1-w^*)^2 \frac{s_L^2}{(1-c)n}.$$

Although all users' data is used here, weighting the estimates by w^* induces excess sampling error $\frac{(w^*-c)^2}{c(1-c)n}\sigma^2$, and the privacy error $w^{*2}s_T^2 + \frac{(1-w^*)^2}{(1-c)n}s_L^2$ is the weighted combination of the groups' privacy errors.

Now we compute and analyze the relative improvement in MSE of the KVH estimator over the best MSE of either the TCM-Only estimator or the Full-LM estimator.

Theorem 4.3. *The relative improvement of $\tilde{\mu}_{KVH}$ over the better of $\tilde{\mu}_T$ and $\tilde{\mu}_F$ is:*

$$R(\mathcal{E}_{KVH}) = \gamma \cdot \begin{cases} \frac{1-c}{cn}\sigma^2 + s_T^2 & \text{if } c > c_{crit} \wedge n \leq n_{crit} \\ \frac{s_L^2}{n} & \text{otherwise} \end{cases},$$

where $\gamma = \frac{(1-c)\sigma^2 s_L^2 + cn(c\sigma^2 + s_L^2)s_T^2}{n(\sigma^2 + cs_L^2) + (1-c)cn^2 s_T^2}$ and c_{crit} and n_{crit} are as defined in Lemma 3.6.

Proof. Direct application of Lemmas 3.2, 3.4, and 4.2 to Definition 3.7. \square

Algebraic analysis of this relative improvement reveals that $R(\mathcal{E}_{KVH}) > 1$ when the number of TCM users is less than s_L^2/s_T^2 . For the common DP mechanisms that apply 0-mean additive noise, this is trivially satisfied. For instance, when adding ϵ -DP Laplace noise, $s_L^2/s_T^2 = c^2 n^2 \geq cn = |T|$. Moreover, although $R(\mathcal{E}_{KVH})$ is theoretically unbounded, using the ϵ -DP Laplace mechanism in the high-privacy regime ($\epsilon \leq 1$) enables a tight characterization of the maximum possible relative improvement.

Corollary 4.4. *The maximum relative utility of $\tilde{\mu}_{KVH}$ when using the Laplace mechanism in the high-privacy regime is bounded as:*

$$17/8 \leq \max_{\substack{\epsilon \leq 1 \\ c, n, m, \sigma}} R(\mathcal{E}_{KVH}) \leq 16/7.$$

Proof. See Appendix B.

Empirical Evaluation of $R(\mathcal{E}_{KVH})$

To better understand what improvements one can expect from $\tilde{\mu}_{KVH}$ in practical applications, we empirically evaluate $R(\mathcal{E}_{KVH})$ using the ϵ -DP Laplace mechanism in the context of various datasets. Note that although the hybrid estimator's performance is dependent on the data distribution only through σ , n , and m , we use datasets to realistically motivate these values.

In Figure 1, we use three synthetic datasets from the Beta(α, β) distribution: Beta(10, 10), Beta(1, 1), and

Beta(0.1, 0.1). These symmetric distributions are chosen to induce different σ values – low ($\sigma \approx 0.109$), medium ($\sigma \approx 0.289$), and high ($\sigma \approx 0.456$). For each distribution, $R(\mathcal{E}_{KVH})$ is plotted across $n \in [10^3, 10^5]$, $c \in \{0.5\%, 5\%\}$, and $\epsilon \in \{0.1, 1\}$. Since the Beta distributions are supported on the interval $[0, 1]$, we let $m = 1$. Figures 1b,c,d show that in these settings, $R(\mathcal{E}_{KVH})$ is lower-bounded by 1 and none are much larger than 2 – matching our theoretical analysis. Observe that the “peaking” behavior of some curves is caused by the the n_{crit} and c_{crit} values being surpassed, which corresponds to the TCM group's data beginning to outperform the LM group's data in terms of MSE. The curves which don't peak either have trivially surpassed the critical values (i.e., $n_{crit} < 1$ with $c > c_{crit}$) or have $c < c_{crit}$; importantly, they don't change behavior at some n not shown in the figures.

In Figure 2, we use a real-world dataset of salaries of $n = 252,540$ employees in the University of California system in 2010 [49]. This dataset was chosen due to its relatively high asymmetry, with a maximum salary of $m \approx 2,349,033$ and standard deviation of $\sigma \approx 53,254$ (both assumed to be known). As σ , n , and m are determined by the dataset, we examine the $R(\mathcal{E}_{KVH})$ values across a large space of the remaining free parameters: $c \in [0.1\%, 10\%]$ and $\epsilon \in [0.1, 10]$. We see the relative improvement peak just above 2 in the high-privacy regime, with this maximum improvement continuing into the low-privacy regime.

5 Homogeneous, Unknown-Variance Setting

In this section, we design a different hybrid estimator for the homogeneous setting, now applied to the case where the variance σ^2 of the data is not known. This is a more realistic setting, as an analyst with no knowledge of the distribution's mean typically also doesn't have knowledge of its variance.

The KVH estimator was able to use knowledge of the variance to weigh the estimates of the two groups so that the trade-off of excess sampling error and privacy error was optimally balanced. In this unknown-variance case, determining the optimal weighting is no longer viable. Nevertheless, we can heuristically choose a weighting which may (or may not) perform well depending on the underlying distribution. Thus, we propose a heuristic weighting choice for combining the groups' estimates and analyze it theoretically and empirically.

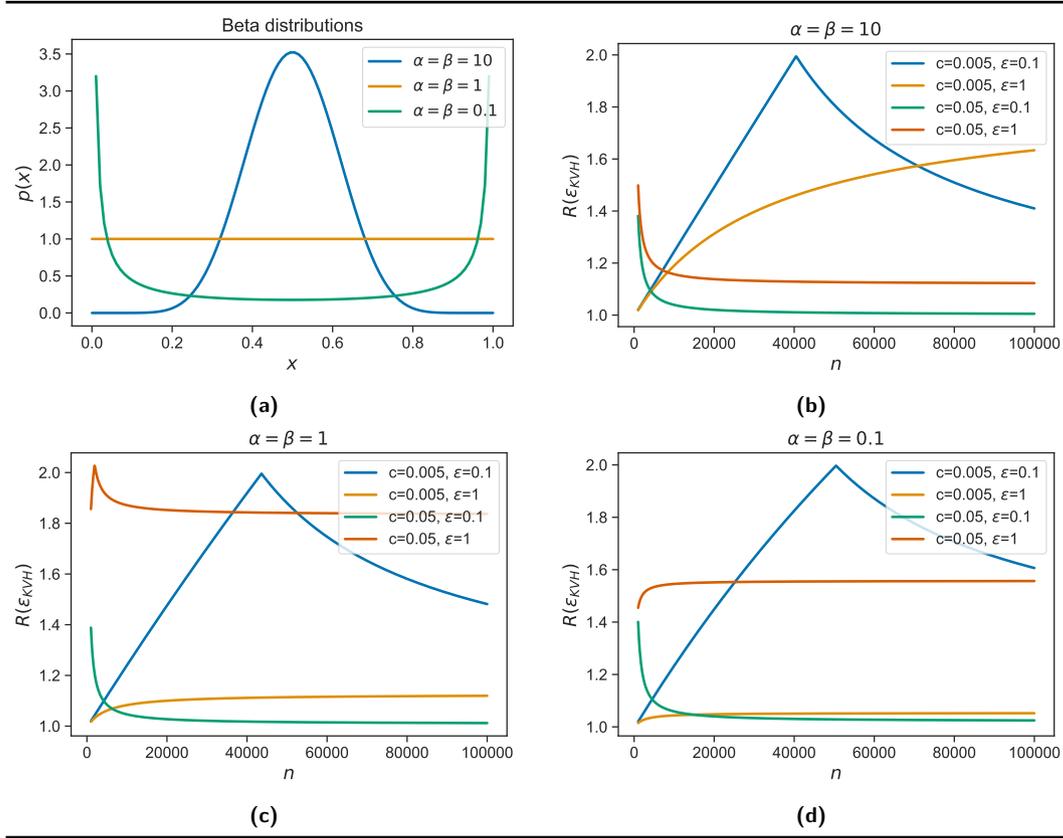


Fig. 1. (a) Probability density functions of $\text{Beta}(\alpha, \beta)$ distributions for various α, β values. (b,c,d) The relative improvement $R(\mathcal{E}_{KVH})$ for each Beta distribution across a range of n values, for various c and ϵ values.

Before detailing this estimator, we first discuss a useful, but weaker, measure of relative improvement for this unknown-variance case.

Utility Over At Least One Baseline

Ideally, estimators would have $R(\mathcal{E}) \geq 1$ for all parameters. If the regions can be computed where each baseline estimator has the best MSE, then a hybrid estimator can be designed to use this knowledge to trivially ensure $R(\mathcal{E}) \geq 1$. However, depending on the setting (such as when variance is unknown), determining these regions precisely may not be feasible. In these cases, we want to at least ensure that the hybrid estimator is never performing worse than both baselines, and do so by defining the following measure of relative improvement.

Definition 5.1. *The relative improvement of an estimator \mathcal{E} over the worst baseline estimator is:*

$$r(\mathcal{E}) = \frac{\max\{\mathcal{E}_T, \mathcal{E}_F\}}{\mathcal{E}}.$$

Our characterization of the critical values in Lemma 3.6 enables $r(\epsilon)$ to be re-written as follows.

Definition 5.2. *The relative improvement of an estimator with MSE \mathcal{E} over the worst baseline estimator is:*

$$r(\mathcal{E}) = \frac{1}{\mathcal{E}} \cdot \begin{cases} \frac{s_L^2}{n} & \text{if } c > c_{crit} \wedge n \leq n_{crit} \\ \frac{1-c}{cn} \sigma^2 + s_T^2 & \text{otherwise} \end{cases}$$

We remark that although any “reasonable” hybrid estimator should satisfy $r(\mathcal{E}) \geq 1$, this criteria is not automatically satisfied. Even among the family of hybrid estimators from Definition 3.12, there exist estimators which have $r(\mathcal{E}_H(w)) < 1$ in some regimes. Concretely, consider an arbitrary constant as the weight; e.g., $w = 0.001$. Using the parameters from experiments ($m = 1, \sigma = 1/6, c = 0.01, \epsilon = 0.1$), we have $r(\mathcal{E}_H(0.001)) < 1$ for $n \geq 10,058$. Thus, estimators must be designed carefully to maximize utility and, at the very least, ensure $r(\mathcal{E}) > 1$ everywhere.

We now propose and analyze a hybrid estimator with a heuristically-chosen weighting that is based on the amount of privacy noise each group adds. However,

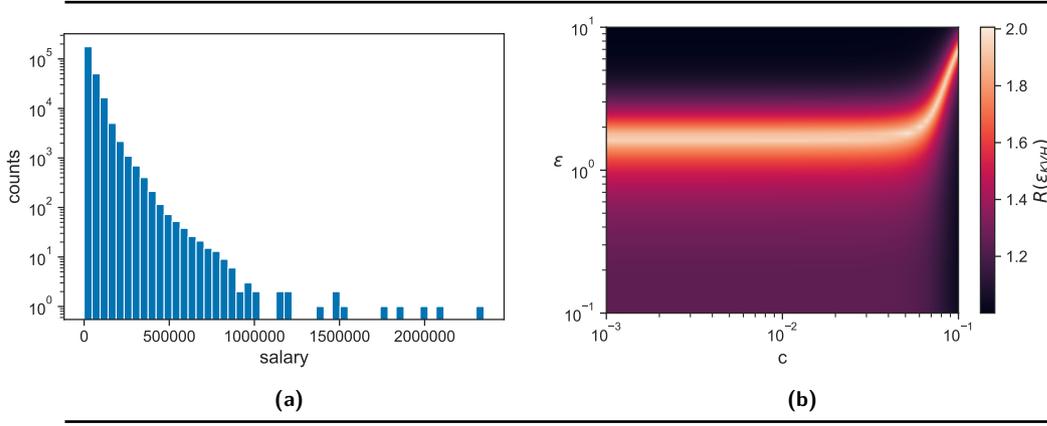


Fig. 2. (a) Distribution of salaries of UC employees. (b) The relative improvement $R(\mathcal{E}_{KVH})$ across a range of c and ϵ values.

we first remark that we additionally investigated a naïve weighting heuristic, which performs the same weighting as the non-private benchmark estimator: weight the estimates based purely on the group size (i.e., $w = c$). Our empirical evaluations showed that for practical parameters, this estimator is typically inferior to the estimator we’re about to discuss. Thus, we have omitted it from this presentation for brevity.

PWH Estimator

We choose this heuristic weighting by considering only the induced privacy error of each groups’ estimate. Thus, we refer to this as the “privacy-weighted hybrid” (PWH) estimator. Note that this weighting seeks solely to optimally balance privacy error between the groups, and therefore ignores the induced excess sampling error. Explicitly, from $\mathcal{E}_H(w)$ of Lemma 3.13 applied to the homogeneous setting, this weighting corresponds to choosing w to minimize $w^2 s_T^2 + (1-w)^2 \frac{s_L^2}{(1-c)n}$, stated in the following definition.

Definition 5.3. *The privacy-weighted hybrid estimator is:*

$$\tilde{\mu}_{PWH} = w_{PWH} \tilde{\mu}_T + (1 - w_{PWH}) \tilde{\mu}_L,$$

$$\text{where } w_{PWH} = \frac{s_L^2}{s_L^2 + (1-c)ns_T^2}$$

Lemma 5.4. *$\tilde{\mu}_{PWH}$ has expected squared error:*

$$\mathcal{E}_{PWH} = \frac{(1-c)cn^2s_T^4(c\sigma^2 + s_L^2) + cns_L^2s_T^2(2(c-1)\sigma^2 + s_L^2) + (1-c)\sigma^2s_L^4}{cn(s_L^2 + (1-c)ns_T^2)^2}.$$

This estimator has a mixture of both excess sampling error and privacy error. Since the privacy error was directly optimized, we expect this estimator to do well

when the data variance σ^2 is small, as this will naturally induce small excess sampling error.

Now we are able to discuss the relative improvement of the PWH estimator over the baselines.

Theorem 5.5. *The relative improvements of the PWH estimator $\tilde{\mu}_{PWH}$ over $\tilde{\mu}_T$ and $\tilde{\mu}_F$ are:*

$$R(\mathcal{E}_{PWH}) = \gamma \cdot \begin{cases} \frac{1-c}{cn}\sigma^2 + s_T^2 & \text{if } c > c_{crit} \wedge n \leq n_{crit} \\ \frac{s_L^2}{n} & \text{otherwise} \end{cases},$$

$$r(\mathcal{E}_{PWH}) = \gamma \cdot \begin{cases} \frac{s_L^2}{n} & \text{if } c > c_{crit} \wedge n \leq n_{crit} \\ \frac{1-c}{cn}\sigma^2 + s_T^2 & \text{otherwise} \end{cases},$$

where $\gamma = \frac{cn(s_L^2 + (1-c)ns_T^2)^2}{(1-c)cn^2s_T^4(c\sigma^2 + s_L^2) + cns_L^2s_T^2(2(c-1)\sigma^2 + s_L^2) + (1-c)\sigma^2s_L^4}$ and c_{crit} and n_{crit} are as defined in Definition 3.7.

Proof. Direct application of Lemmas 3.2, 3.4, and 5.4 to: Definition 3.7 for $R(\mathcal{E}_{PWH})$, and Definition 5.2 for $r(\mathcal{E}_{PWH})$. \square

With the generic noise-addition privacy mechanisms, algebraic analysis of the weaker relative improvement measure reveals $r(\mathcal{E}_{PWH}) > 1$ unconditionally. However, the regions where $R(\mathcal{E}_{PWH})$ is greater than 1 are difficult to obtain analytically with these generic mechanisms. By restricting our attention to the Laplace mechanism, we find that $R(\mathcal{E}_{PWH}) > 1$ is satisfied under certain conditions. The first is a “low relative privacy” regime where $\epsilon \geq \frac{\sqrt{2m}}{\sigma}$; that is, once ϵ is large enough, we have $R(\mathcal{E}_{PWH}) > 1$. For ϵ under this threshold, achieving $R(\mathcal{E}_{PWH}) > 1$ requires the following conditions on c and n : either $c \leq \frac{\epsilon^2\sigma^2}{2m^2}$, or $c > \frac{\epsilon^2\sigma^2}{2m^2} \wedge n < \frac{2m^2(1+c)}{c(2cm^2 - \epsilon^2\sigma^2)}$.

Empirical Evaluation of $R(\mathcal{E}_{PWH})$ and $r(\mathcal{E}_{PWH})$

Here, we perform an empirical evaluation analogous to that done in Section 4. Figure 3 presents $R(\mathcal{E}_{PWH})$ (top row) and $r(\mathcal{E}_{PWH})$ (bottom row) using the same Beta distributions and parameters ($n \in [10^3, 10^5]$, $c \in \{0.5\%, 5\%\}$, and $\epsilon \in \{0.1, 1\}$). We find that there are many regions where $R(\mathcal{E}_{PWH})$ achieves a value of just greater than 1, and some regions where it achieves values competitive with the KVH estimator. Unsurprisingly, since this weighting is chosen without accounting for the variance, there are also clear regions where the $R(\mathcal{E}_{PWH})$ is noticeably less than 1. Even in the regions where $R(\mathcal{E}_{PWH})$ is low, the $r(\mathcal{E}_{PWH})$ values in the bottom row often show that the PWH estimator significantly improves over the worse of the two baseline estimators. An empirical evaluation of this estimator on the UC salaries dataset can be found in Appendix C.

6 Heterogeneous Setting

In this section, we examine the effects of the groups' distributions diverging on the quality of our estimators. This is motivated by the fact that the hybrid model allows users to self-partition based on their trust preferences. Such self-partitioning may cause the groups' distributions to be different. For instance, since the TCM users have similar trust preferences, their data may also be more similar than the LM users'. This could manifest as variance-skewness between the groups. Alternatively, the TCM users may have fundamentally different data than the LM users, which would manifest as mean-skewness between the groups. Thus, we examine the case where the group means are the same but their variances are different, as well as the case where the group means are different but their variances are the same. To understand these skewness effects, we empirically evaluate $R(\mathcal{E}_{KVH})$ ¹.

Although the heterogeneous setting is more general and complex, we can still derive the optimal weighting for the KVH estimator analogously to homogeneous KVH weighting of Definition 4.1.

¹ We also performed the same empirical evaluation with the unknown-variance PWH estimator. The results were nearly identical to the KVH estimator's, and the conclusions were the same. Thus, we omitted them for brevity.

Definition 6.1. *The known-variance hybrid estimator in the heterogeneous setting is:*

$$\tilde{\mu}_{KVH} = w^* \tilde{\mu}_T + (1 - w^*) \tilde{\mu}_L,$$

$$\text{where } w^* = \frac{c(s_L^2 + c\sigma_L^2 + (1-c)(n(\mu_L - \mu)(\mu_L - \mu_T) + \sigma_T^2))}{cs_L^2 + (1-c)cn((\mu_L - \mu_T)^2 + s_T^2) + c\sigma_L^2 + (1-c)\sigma_T^2}$$

Variance-Skewness

Here, we examine the case where $\mu_T = \mu_L$ but $\sigma_T^2 \neq \sigma_L^2$. This reduces the KVH estimator's weighting to $w^* = \frac{c(s_L^2 + c\sigma_L^2 + (1-c)\sigma_T^2)}{cs_L^2 + (1-c)cn s_T^2 + c\sigma_L^2 + (1-c)\sigma_T^2}$. To gain insight into the effect of variance-skewness, we recall two Beta distributions previously used in our empirical evaluations: the low-variance Beta(10,10) distribution ($\sigma = 0.109$) and the high-variance Beta(0.1, 0.1) distribution ($\sigma = 0.456$). We evaluate $R(\mathcal{E}_{KVH})$ in two scenarios: when the TCM group has data drawn from the low-variance distribution but the LM group has data drawn from the high-variance distribution, and vice versa. Figure 4 gives the results across the same range of n , c , and ϵ values as used in previous experiments.

The similarities between Figure 4 and Figure 1 demonstrate that our estimator is robust to deviations in the LM group's variance. For example, Figure 1b shows $R(\mathcal{E}_{KVH})$ when all the data is from the low-variance distribution; that figure nearly exactly matches Figure 4a despite the fact that most of the data is now from the LM group's high-variance distribution. As this applies to both of Figure 1's graphs, it is clear that the relative improvement heavily depends on the variance of the TCM group, regardless of whether the LM group had the low- or high-variance data. In fact, in both graphs, the difference in relative improvement from the homogeneous case with variance σ^2 to the heterogeneous case where only the TCM group has variance $\sigma_T^2 = \sigma^2$ does not vary by more than ± 0.1 , and, typically, varies by less than ± 0.01 .

Mean-Skewness

Here, we examine the case where $\mu_T \neq \mu_L$ but $\sigma_T^2 = \sigma_L^2$. This reduces the KVH estimator's weighting to $w^* = \frac{c(s_T^2 + (1-c)n(\mu_L - \mu)(\mu_L - \mu_T) + \sigma^2)}{cs_T^2 + (1-c)cn((\mu_L - \mu_T)^2 + s_T^2) + \sigma^2}$. Importantly, this expression depends on the curator's knowledge of μ_T and μ_L – an unreasonable requirement, since the curator's overarching goal is to learn the mean from the user data. For applications where the groups' means are assumed to be different, computing separate estimates of each group's mean in their respective trust models would likely be more useful than a joint estimate. Thus, we instead explore mean-skewness from the

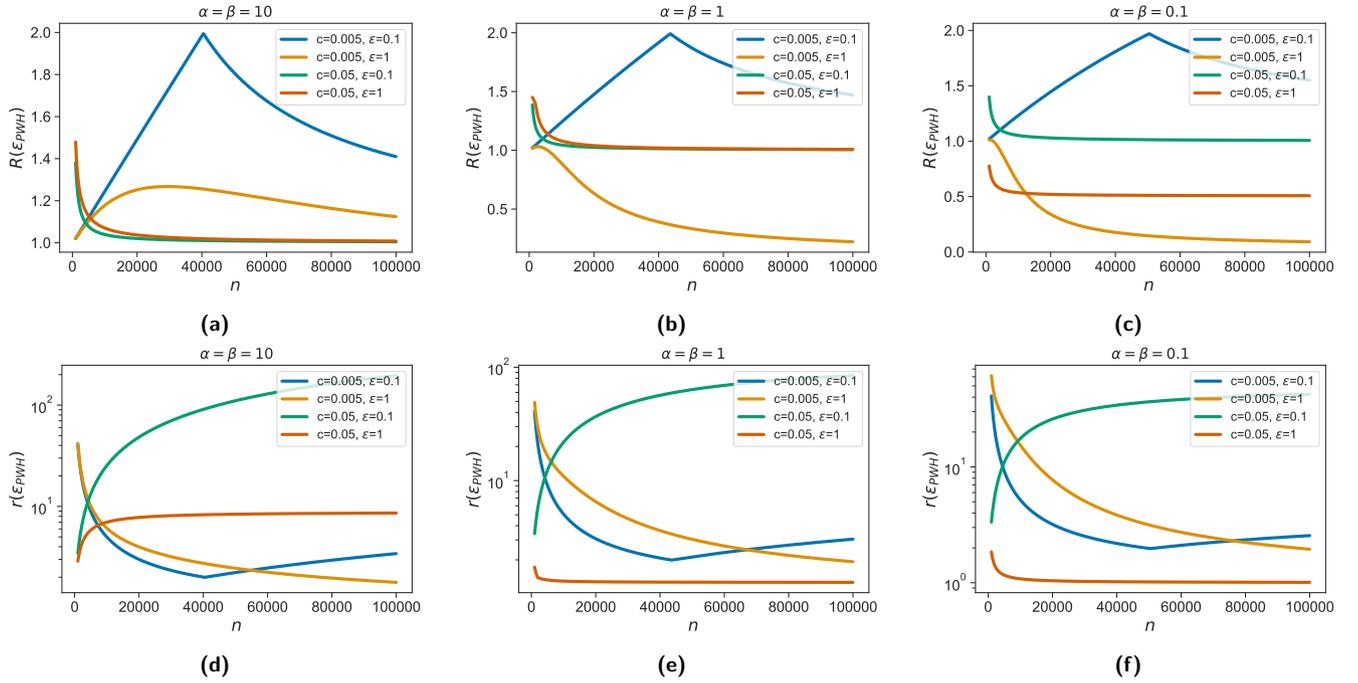


Fig. 3. Across a range of n values, for various c and ϵ values for each Beta distribution (plotted in Figure 1a): (a,b,c) shows $R(\mathcal{E}_{PWH})$ values and (d,e,f) shows $r(\mathcal{E}_{PWH})$ values.

point of view of a curator who *mistakenly believes* they are operating in the homogeneous setting, and thus uses the homogeneous weighting from Definition 4.1. This is useful in practice, as it demonstrates how a curator can use our analytical expressions for their specific problem instance to understand how utility is affected by misspecified assumptions about user data.

To analyze this case, we set up the following experiment, displayed in Figure 5. At the high level, we start with the control for the experiments: set both groups to the same distribution $\mathcal{D}_T = \mathcal{D}_L$ and obtain $R(\mathcal{E}_{KVH})$. Next, we retain the distributional shape for both groups, but shift them in opposite directions; e.g., $\mathcal{D}_T - t, \mathcal{D}_L + t$ for some t . We obtain the new $R(\mathcal{E}_{KVH})$ values under these distributions, and compare against the un-shifted results. For clarity, we denote the relative improvement on the t -shifted distribution as $R^t(\mathcal{E}_{KVH})$.

We expect that as the divergence in means t increases, the relative utility of our hybrid estimator will decrease. To test this hypothesis concretely, we use the medium-variance Beta(1,1) distribution ($\sigma = 0.289$) from our previous empirical evaluations as the experiment’s base distribution. We center this distribution at 1 without rescaling, inducing support on $[0.5, 1.5]$. Then we set both \mathcal{D}_T and \mathcal{D}_L to this distribution, and obtain $R^0(\mathcal{E}_{KVH})$ on it (Figure 5ab). Next, we add a small shift of $t = 0.25$ to each of the groups’ distributions in

opposite directions; i.e., $\mathcal{D}_T - 0.25$ and $\mathcal{D}_L + 0.25$, so that so that $|\mu_T - \mu_L| = 0.5$. These distributions, along with the corresponding $R^{0.25}(\mathcal{E}_{KVH})$ results, are shown in the second column of Figure 5. Finally, the third column of Figure 5 shows the analogous distributions and results when a large shift of $t = 0.5$ is added so that $|\mu_T - \mu_L| = 1$.² Unsurprisingly, these results depict a clear negative impact on the relative improvement as the means diverge, showing that our estimator is sensitive to skewness in the groups’ means.

7 Hybrid Estimator Applications

In this section, we demonstrate how more complex non-hybrid algorithms can be easily extended into the hybrid model using our hybrid estimator as a mean estimation primitive. In particular, we implement a hybrid variant of the classic DP K -means algorithm [23] using the

² One caveat to these shifts is that as the data distribution becomes wider, the noise required to ensure DP must increase. Since we are interested in the effect of mean-skewness here, and not the effect of distribution-width, we conservatively fix $m = 2$ for all experiments. That is, the same level of noise is used across shift-amounts, even if less noise may have sufficed to ensure DP.

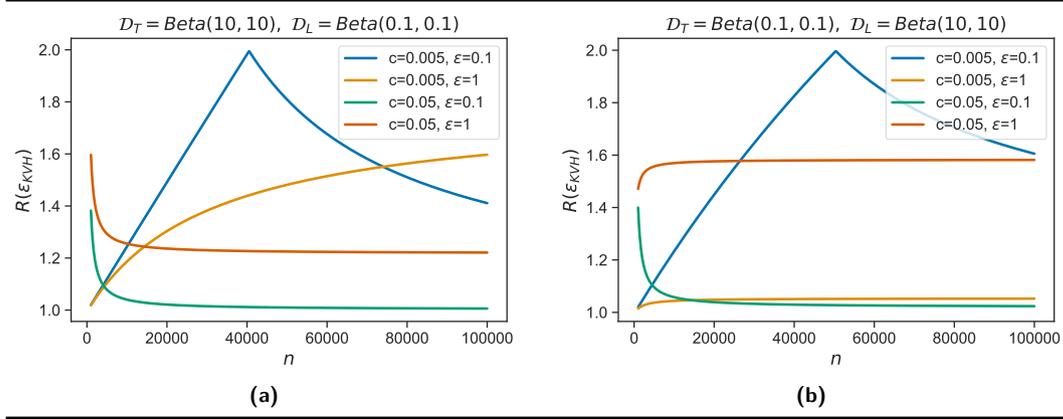


Fig. 4. The relative improvement $R(\mathcal{E}_{KVH})$ values when: (a) the TCM group has low-variance data but the LM group has high, and (b) when the TCM group has high-variance data but the LM group has low.

PWH hybrid estimator as a sub-component, then empirically evaluate its effectiveness.

The K -means problem is to partition n d -dimensional real-valued observations x_1, \dots, x_n into K clusters C_1, \dots, C_K such that the within-cluster sum of squares (WCSS) is minimized. Denoting μ_k as the center of cluster C_k , this is formally:

$$\arg \min_{C_1, \dots, C_K} \sum_{k=1}^K \sum_{x \in C_k} \|x - \mu_k\|^2.$$

This problem is NP-hard, and thus heuristic algorithms are generally used. The classic DP algorithm for this problem was designed for the TCM and analyzed in [23]. This algorithm partitions the total privacy budget across τ iterations, and each iteration refines the estimates of the clusters' centers. Each iterative refinement assigns the observations to their nearest cluster, then updates each cluster's center to the mean of all points within it while carefully applying Laplace noise.

We extend this algorithm to the LM in a simple way. First, LM users expend a portion of their privacy budget reporting their data to the curator with Laplace noise. The curator uses their data analogously to the TCM case, except that in each iteration, LM users use a portion of their privacy budget to report the nearest cluster to them using randomized response – this reduces bias in the cluster centers, relative to computing the nearest cluster directly based on their already-reported data.

Other DP K -means algorithms exist in both the TCM [4, 8, 43, 46, 48, 53] and LM [47, 52, 54, 57] which improve on our two non-hybrid K -means algorithms. However, the purpose of this section is to demonstrate how our hybrid estimator can be effectively leveraged in more complex applications. Thus, we present our hybrid K -means algorithm in Appendix E, which combines

our simpler TCM and LM algorithms in the following straightforward way. Each separate algorithm performs its iterative refinement as previously described. Then, at the end of each iteration, the TCM and LM cluster center estimates are combined using the PWH estimator on each dimension.

We evaluate the hybrid algorithm in the following experiment, showing that it automatically achieves WCSS on-par with the best baseline. The baselines here, analogous to our estimators' TCM-Only and Full-LM MSE baselines, are: the WCSS of the TCM variant using only TCM data, and the WCSS of the LM variant using all data. The dataset used for evaluations is shown in Figure 6a: 4 clusters of 2-dimensional spherical Gaussian data with scale $\sigma \approx 0.028$ and 40,000 points per cluster. In Figure 6bc, across a range of total iterations τ and fractions of TCM users 0.1% and 1%, we evaluate the mean WCSS values of each model's algorithm with 364 trials. The privacy budget for each algorithm is $\epsilon = 7$; this relatively high budget is necessitated for the TCM and LM algorithms to achieve acceptable practical utility. The regimes where each non-hybrid algorithm is better than the other is unclear a priori, and the results here show one example of each. By simply combining the two using our hybrid estimator, the hybrid algorithm is able to maintain a WCSS approximately equal to the better of the two.

8 Privacy Amplification Via Inter-group Interaction

The benefit or necessity of inter-group interaction in the hybrid model is an active area of research. Aven

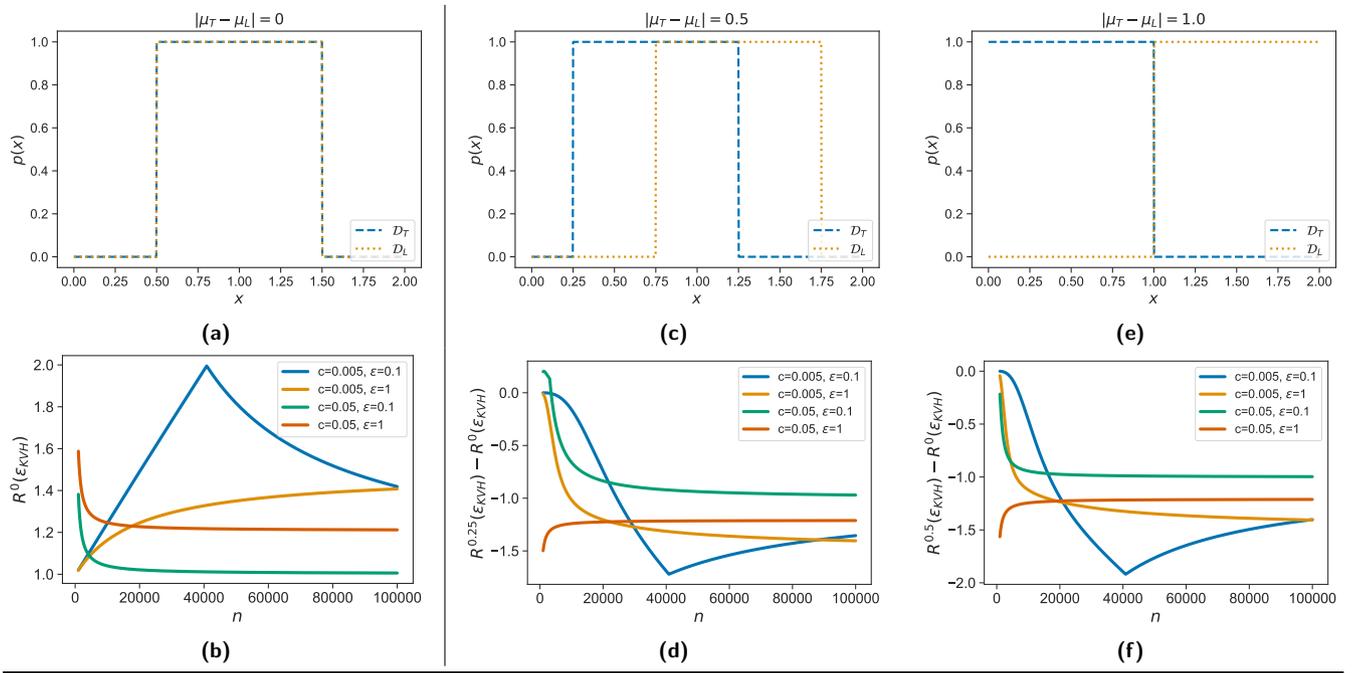


Fig. 5. (Left column) Initial data distributions with no mean shift, and the KVH estimator’s corresponding relative improvement. Small (middle column) and large (right column) mean shifts of the initial data distribution with $t = 0.25$ and $t = 0.5$ respectively, along with the KVH estimator’s corresponding change in relative improvement.

et al. [3] showed experimentally that good utility is achievable by intelligently utilizing inter-group interaction. In this work, we’ve shown mathematically that we can guarantee good utility for mean estimation with no inter-group interactivity. However, both works only focus on inter-group interactivity’s effect on a mechanism’s *utility* – neither consider its effect on *privacy*. Each group is assumed to independently guarantee privacy, without considering how subsequent interaction and processing by the curator may effect the DP guarantee. The post-processing property of DP ensures that such interaction and processing won’t degrade privacy, but the question of whether it improves privacy was unstudied. It is precisely this effect on privacy that we introduce and examine in this section. We find that for our hybrid mean estimators, the privacy guarantee against certain adversaries can be significantly improved.

We are specifically interested in users’ privacy against adversaries who can view the output of the curator’s computation (i.e., output-viewing adversaries). This is the classic adversary model that the TCM protects against. The LM protects against a larger class of adversaries: the output-viewing adversaries, as well as against the curator itself. However, the LM’s singular DP guarantee doesn’t distinguish between these adversary types. In the hybrid model, each groups’ DP

guarantee may be overly-conservative against output-viewing adversaries since it doesn’t account for the curator’s joint processing of the LM users’ reports – which each include their own privacy noise – in conjunction with the TCM group’s privacy noise. Thus, we investigate users’ DP guarantee against output-viewing adversaries as a result of: 1) the combined privacy noise from both groups, in conjunction with 2) the inter-group interaction strategy of the curator. We show that these two components together can serve to amplify users’ privacy against this adversary class. This provides a two-tier DP guarantee for LM users – their standard DP guarantee against the curator, and an improved guarantee against output-viewing adversaries – and an improved DP guarantee for TCM users. To make this concrete, we first analyze our hybrid estimator family and show how its non-interactive strategy can amplify privacy. We then describe why BLENDER’s [3] interaction strategy does not provide such amplification. Together, these examples highlight the value of looking at the effects of inter-group interaction not only on utility, but also on privacy.

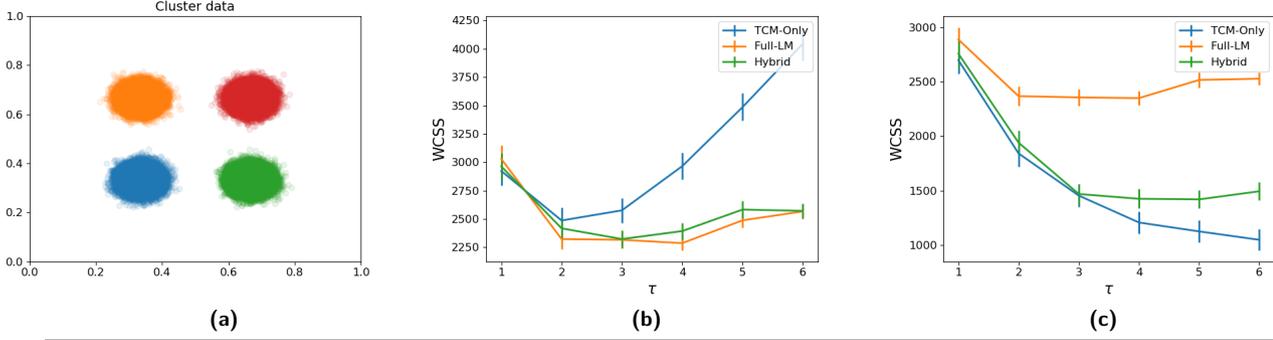


Fig. 6. (a) Clustering dataset with 4 clusters of 2-d spherical Gaussians with $\sigma \approx 0.028$ and 40,000 points per cluster. (b,c) WCSS values of each model’s algorithm across a range of total iterations τ , 0.1% and 1% fractions of TCM users respectively, and $\epsilon = 7$.

Hybrid Mean Estimator Amplification

Recall that the hybrid estimator family from Definition 3.12 utilizes no inter-group interaction – i.e., the curator only outputs once: after it has received all the LM users reports, computed both groups’ estimates, then combined them. For adversaries that can only view the output of this curator, the combined noise from *all* the LM users and the TCM group can serve to improve the DP guarantee. To see this, we re-write the estimator as

$$\begin{aligned} \tilde{\mu}_H(w) = & \underbrace{\left(\frac{w}{cn} \sum_{i \in T} x_i + \frac{1-w}{(1-c)n} \sum_{i \in L} x_i \right)}_{\text{non-private hybrid mean estimator}} \\ & + \underbrace{\left(wY_T + \frac{1-w}{(1-c)n} \sum_{i \in L} Y_{L,i} \right)}_{\text{joint privacy noise}}. \end{aligned}$$

Thus, this joint privacy noise is providing some DP guarantee for the mechanism as a whole, rather than individual noises protecting the individual groups.

There is one caveat: the TCM users’ noise is provided by the curator and never revealed to them, but the LM users each provide their own noise. DP requires that the privacy noise not be known to an adversary; any noise that is known cannot be considered towards the DP guarantee. Here, we assume LM users are semi-honest – i.e., they apply the specified mechanism properly to their data, but they know the privacy noise they add. Thus, LM user i ’s knowledge of their own privacy noise weakens the the joint noise term by an additive $\frac{1-w}{(1-c)n} Y_{L,i}$ amount. Furthermore, they may choose to form coalitions with other users and share this knowledge to adversarially weaken the joint privacy noise term. The largest such coalition, denoted as A , reduces the joint privacy noise by $\frac{1-w}{(1-c)n} \sum_{i \in A} Y_{L,i}$. Excluding

the largest such coalition’s noise enables the remaining joint privacy noise to be analyzed for a DP guarantee.

The DP guarantee from the remaining joint noise depends on the privacy mechanisms used by the TCM group and each LM user. For instance, the ϵ -DP Laplace mechanism would yield a joint noise term which guarantees ϵ' -DP where $\epsilon' = \epsilon -$ i.e., it wouldn’t enable any privacy amplification (see proof in Appendix D). Alternatively, consider the Gaussian mechanism, where the curator adds $Y_T \sim \mathcal{N}(0, s_T^2)$ and each LM user i adds $Y_{L,i} \sim \mathcal{N}(0, s_L^2)$, where s_T^2 and s_L^2 are calibrated to ensure (ϵ, δ) -DP for both groups. Now, analyzing this joint noise provides the following amplified DP guarantee against output-viewing adversaries.

Theorem 8.1. *Assume the curator adds Gaussian noise of variance s_T^2 to provide an (ϵ, δ) -DP guarantee for the TCM group, and that each LM user adds Gaussian noise of variance s_L^2 to provide an (ϵ, δ) -DP guarantee for themselves. Furthermore, assume that the largest adversarial coalition is of size $|A|$. Define*

$$s'^2 = w^2 s_T^2 + \left(\frac{1-w}{(1-c)n} \right)^2 |L \setminus A| s_L^2.$$

The users’ ϵ' -DP guarantee against output-viewing adversaries is given by:

$$\epsilon' = \frac{\sqrt{2 \ln(1.25/\delta)} m}{ns'} \cdot \begin{cases} \frac{w}{c}, & w \leq c \\ \frac{1-w}{1-c}, & \text{otherwise} \end{cases}$$

Proof. See Appendix D.

For practical applications, even with a moderate fraction of adversarial LM users, this amplification can be significant. To make this concrete, consider the UC salary dataset used in the previous experiments. Suppose we compute the KVH estimator with each group

using the Gaussian mechanism with $\epsilon = 1$ and $\delta = 10^{-7}$. In Figure 7, we plot the users’ amplified ϵ' value across $c \in [0.1\%, 10\%]$ as well as across the fraction of LM users assumed to be adversarial.

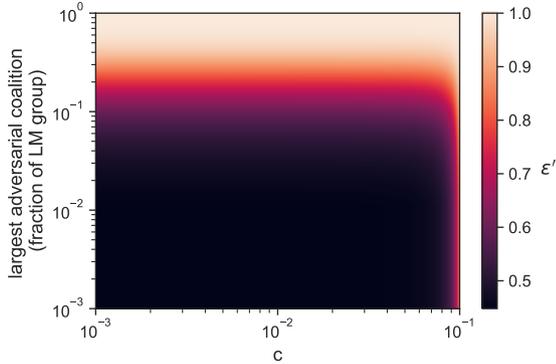


Fig. 7. The amplified (ϵ', δ) -DP guarantee when the $(1, 10^{-7})$ -DP Gaussian mechanism is used in the KVH estimator.

BLENDER Amplification

Now that we’ve shown how a *lack* of inter-group interaction can facilitate privacy amplification, we turn our focus to BLENDER’s inter-group interactions strategy for improving utility. As described in Section 2.3, BLENDER takes advantage of the TCM group by having it identify the heavy hitters. The TCM group then passes the identified heavy hitters on to the LM users, who perform frequency estimation. The curator then combines the LM users’ reports and outputs the heavy hitters along with their frequencies.

One might be tempted to analyze this final output for an amplified DP guarantee. However, the initial output of the curator – the privatized list of heavy hitters from the TCM group – has already been released to all LM users. Unless *all* LM users are non-adversarial, the TCM users gain no further benefit from the incorporation of the LM users’ privacy noise. Conversely, the LM users may experience privacy amplification through the combination of their locally-added noise; however, this is solely due to intra-group interaction in the LM, and related topics are currently an active area of research on the LM [5, 6, 16, 19, 26, 33].

9 Conclusions

Under the hybrid model of differential privacy, which permits a mixture of the trusted-curator and local trust models, we explored the problem of accurately estimating the mean of arbitrary distributions with bounded support. We designed a hybrid estimator for the joint mean of both trust models’ users and derived analytically exact finite sample utility expressions – applicable even when the trust models may have different user data distributions.

When the trust models have the same distribution and the curator knows its variance, we proved that our estimator is able to always achieve higher utility than both baseline estimators in the two individual trust models. When the variance is unknown, for many practical parameters, we showed that our hybrid estimator achieves utility nearly as good as when the variance is known. For both cases, we evaluated our estimator on realistic datasets and found that it achieves good utility in practice.

By designing a hybrid variant of the classic differentially private K -means algorithm, we showed how more complex hybrid algorithms can be designed by using the hybrid estimator as a sub-component. Experimentally, we found that this hybrid algorithm automatically achieves utility on-par with the best of its corresponding two non-hybrid algorithms, even though it is unclear a priori when each non-hybrid algorithm is better.

Finally, we introduced a notion of privacy amplification that arises in the hybrid model due to interaction between the two underlying trust models. We derived the privacy amplification that our hybrid estimator provides all users, and discussed when other hybrid mechanisms may fail to achieve amplification.

Acknowledgements

This work was supported by NSF grants No. 1755992, 191615, and 1943584, a VMWare fellowship, and a gift from Google.

The work was partially done while the authors B. A. and A. K. were at the “Data Privacy: Foundations and Applications” Spring 2019 program held at the Simons Institute for the Theory of Computing, UC Berkeley.

References

- [1] M Abramowitz and IA Stegun. *Handbook of Mathematical Functions*. US Government Printing Office, Washington, DC, 7th edition, 1968.
- [2] Jayadev Acharya, Gautam Kamath, Ziteng Sun, and Huanyu Zhang. Inspectre: Privately estimating the unseen. In *International Conference on Machine Learning*, pages 30–39, 2018.
- [3] Brendan Avent, Aleksandra Korolova, David Zeber, Torgeir Hovden, and Benjamin Livshits. BLENDER: Enabling local search with a hybrid differential privacy model. In *26th USENIX Security Symposium (USENIX Security 17)*, pages 747–764. USENIX Association, 2017.
- [4] Maria-Florina Balcan, Travis Dick, Yingyu Liang, Wenlong Mou, and Hongyang Zhang. Differentially private clustering in high-dimensional euclidean spaces. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70, pages 322–331, 2017.
- [5] Victor Balcer and Albert Cheu. Separating local & shuffled differential privacy via histograms. In *1st Conference on Information-Theoretic Cryptography (ITC 2020)*, volume 163 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:14, 2020.
- [6] Borja Balle, James Bell, Adria Gascon, and Kobbi Nissim. Differentially private summation with multi-message shuffling. *arXiv preprint arXiv:1906.09116*, 2019.
- [7] Borja Balle, James Bell, Adrià Gascón, and Kobbi Nissim. The privacy blanket of the shuffle model. In *Annual International Cryptology Conference*, pages 638–667. Springer, 2019.
- [8] Artem Barger and Dan Feldman. k-means for streaming and distributed big sparse data. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 342–350. SIAM, 2016.
- [9] Raef Bassily, Albert Cheu, Shay Moran, Aleksandar Nikolov, Jonathan Ullman, and Zhiwei Steven Wu. Private query release assisted by public data. *arXiv preprint arXiv:2004.10941*, 2020.
- [10] Raef Bassily, Shay Moran, and Noga Alon. Limits of private learning with access to public data. In *Advances in Neural Information Processing Systems*, pages 10342–10352, 2019.
- [11] Raef Bassily and Adam Smith. Local, private, efficient protocols for succinct histograms. In *Proceedings of the Symposium on Theory of Computing (STOC)*, pages 127–135, 2015.
- [12] Amos Beimel, Aleksandra Korolova, Kobbi Nissim, Or Sheffet, and Uri Stemmer. The Power of Synergy in Differential Privacy: Combining a Small Curator with Local Randomizers. In *1st Conference on Information-Theoretic Cryptography (ITC 2020)*, volume 163 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:25, 2020.
- [13] Amos Beimel, Kobbi Nissim, and Uri Stemmer. Private learning and sanitization: Pure vs. approximate differential privacy. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 363–378. Springer, 2013.
- [14] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.
- [15] Sourav Biswas, Yihe Dong, Gautam Kamath, and Jonathan Ullman. Coinpress: Practical private mean and covariance estimation. *arXiv preprint arXiv:2006.06618*, 2020.
- [16] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, pages 441–459. ACM, 2017.
- [17] Avrim Blum, Cynthia Dwork, Frank McSherry, and Kobbi Nissim. Practical privacy: the SuLQ framework. In *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 128–138, 2005.
- [18] Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. In *Advances in Neural Information Processing Systems*, pages 181–191, 2019.
- [19] Albert Cheu, Adam Smith, Jonathan Ullman, David Zeber, and Maxim Zhilyaev. Distributed differential privacy via shuffling. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 375–403. Springer, 2019.
- [20] Wenxin Du, Canyon Foot, Monica Moniot, Andrew Bray, and Adam Groce. Differentially private confidence intervals. *arXiv preprint arXiv:2001.02285*, 2020.
- [21] John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.
- [22] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.
- [23] Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.
- [24] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006.
- [25] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284. Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [26] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2468–2479. SIAM, 2019.
- [27] Giulia Fanti, Vasył Pihur, and Úlfar Erlingsson. Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries. *Proceedings on Privacy Enhancing Technologies (PETS)*, 3:41–61, 2016.
- [28] Vitaly Feldman. Dealing with range anxiety in mean estimation via statistical queries. In *International Conference on Algorithmic Learning Theory*, pages 629–640, 2017.

- [29] Marco Gaboardi, Ryan Rogers, and Or Sheffet. Locally private mean estimation: z -test and tight confidence intervals. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89, pages 2545–2554. PMLR, 2019.
- [30] Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Pure differentially private summation from anonymous messages. *arXiv preprint arXiv:2002.01919*, 2020.
- [31] Badih Ghazi, Noah Golowich, Ravi Kumar, Rasmus Pagh, and Ameya Velingker. On the power of multiple anonymous messages. *arXiv preprint arXiv:1908.11358*, 2019.
- [32] Badih Ghazi, Pasin Manurangsi, Rasmus Pagh, and Ameya Velingker. Private aggregation from fewer anonymous messages. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 798–827. Springer, 2020.
- [33] Badih Ghazi, Rasmus Pagh, and Ameya Velingker. Scalable and differentially private distributed aggregation in the shuffled model. *arXiv preprint arXiv:1906.08320*, 2019.
- [34] Andy Greenberg. Apple’s differential privacy is about collecting your data – but not your data. In *Wired*, June 13, 2016.
- [35] Jihun Hamm, Yingjun Cao, and Mikhail Belkin. Learning privately from multiparty data. In *International Conference on Machine Learning*, pages 555–563, 2016.
- [36] Zhanglong Ji and Charles Elkan. Differential privacy based on importance weighting. *Machine learning*, 93(1):163–183, 2013.
- [37] Matthew Joseph, Janardhan Kulkarni, Jieming Mao, and Steven Z Wu. Locally private gaussian estimation. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2980–2989, 2019.
- [38] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2879–2887, 2014.
- [39] Gautam Kamath, Jerry Li, Vikrant Singhal, and Jonathan Ullman. Privately learning high-dimensional distributions. In Alina Beygelzimer and Daniel Hsu, editors, *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99, pages 1853–1902. PMLR, 2019.
- [40] Gautam Kamath, Or Sheffet, Vikrant Singhal, and Jonathan Ullman. Differentially private algorithms for learning mixtures of separated gaussians. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 168–180, 2019.
- [41] Gautam Kamath, Vikrant Singhal, and Jonathan Ullman. Private mean estimation of heavy-tailed distributions. *arXiv preprint arXiv:2002.09464*, 2020.
- [42] Vishesh Karwa and Salil Vadhan. Finite sample differentially private confidence intervals. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.
- [43] Zhigang Lu and Hong Shen. Differentially private k -means clustering with guaranteed convergence. *arXiv preprint arXiv:2002.01043*, 2020.
- [44] Mary Madden and Lee Rainie. Americans’ attitudes about privacy, security and surveillance. Technical report, Pew Research Center, 2015.
- [45] Chris Merriman. Microsoft reminds privacy-concerned Windows 10 beta testers that they’re volunteers. In *The Inquirer*, <http://www.theinquirer.net/2374302>, Oct 7, 2014.
- [46] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.
- [47] Kobbi Nissim and Uri Stemmer. Clustering algorithms for the centralized and local models. In Firdaus Janoos, Mehryar Mohri, and Karthik Sridharan, editors, *Proceedings of Algorithmic Learning Theory*, volume 83 of *Proceedings of Machine Learning Research*, pages 619–653. PMLR, 07–09 Apr 2018.
- [48] Richard Nock, Raphaël Canyasse, Rokhsana Boreli, and Frank Nielsen. k -variates $++$: more pluses in the k -means $++$. In *International Conference on Machine Learning (ICML)*, pages 145–154, 2016.
- [49] U. of California. 2010 annual report on employee compensation. <https://ucnet.universityofcalifornia.edu/compensation-and-benefits/compensation/index.html>.
- [50] Nicolas Papernot, Martín Abadi, Úlfar Erlingsson, Ian J. Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*, 2017.
- [51] Tiberiu Popoviciu. Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica*, 9:129–145, 1935.
- [52] Uri Stemmer. Locally private k -means clustering. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 548–559. SIAM, 2020.
- [53] Dong Su, Jianneng Cao, Ninghui Li, Elisa Bertino, and Hongxia Jin. Differentially private k -means clustering. In *Proceedings of the sixth ACM conference on data and application security and privacy*, pages 26–37, 2016.
- [54] Lin Sun, Jun Zhao, and Xiaojun Ye. Distributed clustering in the anonymized space with local differential privacy. *arXiv preprint arXiv:1906.11441*, 2019.
- [55] Di Wang, Huanyu Zhang, Marco Gaboardi, and Jinhui Xu. Estimating smooth GLM in non-interactive local differential privacy model with public unlabeled data. *arXiv preprint arXiv:1910.00482*, 2019.
- [56] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.
- [57] Chang Xia, Jingyu Hua, Wei Tong, and Sheng Zhong. Distributed k -means clustering guaranteeing local differential privacy. *Computers & Security*, 90:101699, 2020.
- [58] Sijie Xiong, Anand D Sarwate, and Narayan B Mandayam. Randomized requantization with local differential privacy. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2189–2193. IEEE, 2016.

A Estimator MSE Proofs

Proof of Lemma 3.2

$$\begin{aligned}
\mathbb{V}[\tilde{\mu}_T - \hat{\mu}] &= \mathbb{V}\left[\frac{1}{cn} \sum_{i \in T} x_i + Y_T - \frac{1}{n} \sum_{i \in [n]} x_i\right] \\
&= \mathbb{V}\left[\frac{1}{cn} \sum_{i \in T} x_i - \frac{1}{n} \sum_{i \in T} x_i - \frac{1}{n} \sum_{i \in L} x_i + Y_T\right] \\
&= \frac{(1-c)^2}{cn} \sigma_T^2 + \frac{1-c}{n} \sigma_L^2 + s_T^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{E}_T &= \mathbb{E}[(\tilde{\mu}_T - \hat{\mu})^2] \\
&= \mathbb{V}[\tilde{\mu}_T - \hat{\mu}] + \mathbb{E}[\tilde{\mu}_T - \hat{\mu}]^2 \\
&= \frac{(1-c)^2}{cn} \sigma_T^2 + \frac{1-c}{n} \sigma_L^2 + s_T^2 + (\mu_T - \mu)^2.
\end{aligned}$$

Proof of Lemma 3.4

$$\begin{aligned}
\mathcal{E}_F &= \mathbb{E}[(\tilde{\mu}_F - \hat{\mu})^2] \\
&= \mathbb{V}[\tilde{\mu}_F - \hat{\mu}] + \underbrace{\mathbb{E}[\tilde{\mu}_F - \hat{\mu}]^2}_0 \\
&= \mathbb{V}\left[\frac{1}{n} \sum_{i \in [n]} (x_i + Y_{L,i}) - \frac{1}{n} \sum_{i \in [n]} x_i\right] \\
&= \frac{s_L^2}{n}.
\end{aligned}$$

Proof of Lemma 3.11

$$\begin{aligned}
\mathbb{V}[\tilde{\mu}_L - \hat{\mu}] &= \mathbb{V}\left[\frac{1}{(1-c)n} \sum_{i \in L} (x_i + Y_{L,i}) - \frac{1}{n} \sum_{i \in [n]} x_i\right] \\
&= \mathbb{V}\left[\frac{c}{(1-c)n} \sum_{i \in L} x_i - \frac{1}{n} \sum_{i \in T} x_i + \frac{1}{(1-c)n} \sum_{i \in L} Y_{L,i}\right] \\
&= \frac{c^2}{(1-c)n} \sigma_L^2 + \frac{c}{n} \sigma_T^2 + \frac{1}{(1-c)n} s_L^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{E}_L &= \mathbb{E}[(\tilde{\mu}_L - \hat{\mu})^2] \\
&= \mathbb{V}[\tilde{\mu}_L - \hat{\mu}] + \mathbb{E}[\tilde{\mu}_L - \hat{\mu}]^2 \\
&= \frac{c^2}{(1-c)n} \sigma_L^2 + \frac{c}{n} \sigma_T^2 + \frac{1}{(1-c)n} s_L^2 + (\mu_L - \mu)^2.
\end{aligned}$$

Proof of Lemma 3.13

$$\begin{aligned}
\mathbb{V}[\tilde{\mu}_H(w) - \hat{\mu}] &= \mathbb{V}[w\tilde{\mu}_T + (1-w)\tilde{\mu}_L - \hat{\mu}] \\
&= \mathbb{V}[w\tilde{\mu}_T - c\hat{\mu}_T + (1-w)\tilde{\mu}_L - (1-c)\hat{\mu}_L] \\
&= \frac{(w-c)^2}{cn} \sigma_T^2 + \frac{(w-c)^2}{(1-c)n} \sigma_L^2 + w^2 s_T^2 + \frac{(1-w)^2}{(1-c)n} s_L^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\mathcal{E}_H(w) &= \mathbb{E}[(\tilde{\mu}_H(w) - \hat{\mu})^2] \\
&= \mathbb{V}[\tilde{\mu}_H(w) - \hat{\mu}] + \mathbb{E}[\tilde{\mu}_H(w) - \hat{\mu}]^2 \\
&= \frac{(w-c)^2}{cn} \sigma_T^2 + \frac{(w-c)^2}{(1-c)n} \sigma_L^2 + w^2 s_T^2 + \frac{(1-w)^2}{(1-c)n} s_L^2 \\
&\quad + (w\mu_T + (1-w)\mu_L - \mu)^2.
\end{aligned}$$

B Proof of Corollary 4.4

Note the following for upper-bounding $R(\mathcal{E}_{KVH})$. Popoviciu's inequality [51] states that a random variable bounded in $[a, b]$ has variance at most $(b-a)^2/4$. For our purposes, this ensures $\sigma^2 \leq m^2/4$. For real-world use cases, it is realistic to constrain ϵ to the ‘‘high-privacy’’ regime of $\epsilon \leq 1$. Thus, with $\epsilon \leq 1$ and $\sigma^2 \leq m^2/4$, we have $0 \leq \epsilon^2 \sigma^2 / m^2 \leq 1/4$. Let $y = \epsilon^2 \sigma^2 / m^2$. Now, we upper-bound the improvement ratio as follows.

$$R(\mathcal{E}_{KVH}) \leq \frac{2(2-c)m^2}{2m^2 - (1-c)\epsilon^2 \sigma^2} = \frac{2(2-c)}{2 - (1-c)y} \leq 16/7,$$

where the final inequality stems from constrained maximization across $c \in [0, 1]$ and $y \in [0, 1/4]$ (justified in the above note).

A lower-bound is given by the following concrete instance. Let $m = 1$, $\epsilon = 1$, $\sigma^2 = 1/4$, and $c = \frac{1}{18} \left(1 + \sqrt{\frac{288+n}{n}}\right)$. Then, as $n \rightarrow \infty$, we have that $R(\mathcal{E}_{KVH})$ converges to $17/8$.

C PWH Utility (continued)

Figure 8 presents heatmaps of $R(\mathcal{E}_{PWH})$ and $r(\mathcal{E}_{PWH})$ for the UC salaries dataset across the same parameters as before ($c \in [0.1\%, 10\%]$ and $\epsilon \in [0.1, 10]$). We find that $R(\mathcal{E}_{PWH})$ achieves a value of slightly greater than 1 across a large portion of the space. The results here tell a similar story to that of Figure 3. Most of the space has $R(\mathcal{E}_{PWH})$ values above 1, and even approaching 2 in narrow region. There is also a small region at the large c values where the relative improvement drops below 0.5. The majority of the space has $r(\mathcal{E}_{PWH})$ between

10 and 100, although it includes a region at the high ϵ values where this relative improvement exceeds 1,500.

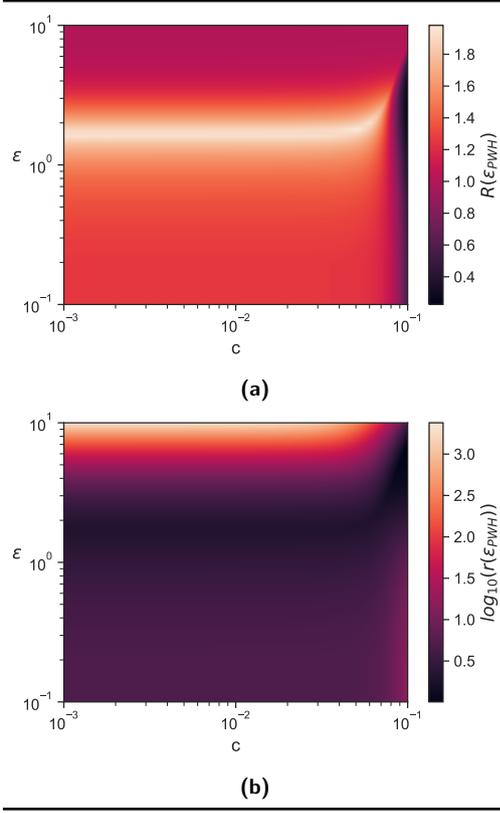


Fig. 8. The relative improvements $R(\mathcal{E}_{PWH})$ (a) and $r(\mathcal{E}_{PWH})$ (b) across a range of c and ϵ values. Note the log scale on (b).

D Privacy Amplification Proofs

ϵ -DP Laplace Mechanism Amplification

Here, we show that the unweighted sum of user's reports, each privatized by their own ϵ -DP Laplace mechanism, only provides a ϵ' -DP joint guarantee of $\epsilon' = \epsilon$. We note that a convex weighting of the terms in this sum yields the same joint guarantee, proving our claim. To formalize this, we define

$$\tilde{S}_n := \sum_{i \in [n]} (x_i + Y_i) = \sum_{i \in [n]} x_i + \sum_{i \in [n]} Y_i$$

where $x_i \in [0, m]$, $Y_i \sim \text{Lap}(b)$, and $b = m/\epsilon$ for each user i . Denote $S_n := \sum_{i \in [n]} x_i$ and $Y := \sum_{i \in [n]} Y_i$. We show here that the joint noise Y provides ϵ' -DP for S_n against output-viewing adversaries, where $\epsilon' = \epsilon$.

First, note that each user at least has the ϵ -DP guarantee via their own privacy noise. Thus, by the post-

processing property of DP, we have $\epsilon' \leq \epsilon$ as a trivial upper-bound. Note that if $\epsilon \leq \epsilon'$ without any adversarial users, then our upper-bound implies $\epsilon \leq \epsilon'$ with an arbitrary number of adversarial users. Therefore, we assume w.l.o.g. that no users are adversarial.

Towards lower-bounding ϵ' , we compute the characteristic function of Y as the product of the characteristic functions of each Y_i . Y 's probability density function, $p_Y(x)$, is then recovered from its characteristic function [14] via the inverse Fourier transform to yield

$$p(x) = \frac{2^{\frac{1}{2}-n}}{\sqrt{\pi} b^{\frac{1}{2}+n} \Gamma(n)} \mathcal{K}_{\frac{1}{2}-n} \left(\frac{|x|}{b} \right) |x|^{n-\frac{1}{2}},$$

where $\overline{\varphi_Y(t)}$ is the complex conjugate of $\varphi_Y(t)$ and $\mathcal{K}(\cdot)$ is the modified Bessel function of the second kind [1]. For ϵ' -DP, noting that $\Delta_1 S_n = m$, we must bound

$$-\epsilon' \leq \max_{k \in [-m, m]} \log \left(\frac{p(x)}{p(x+k)} \right) \leq \epsilon'.$$

Consider the instance where $k = m$ and $x \rightarrow \infty$:

$$\begin{aligned} & \lim_{x \rightarrow \infty} \log \left(\frac{p(x)}{p(x+k)} \right) \\ &= \lim_{x \rightarrow \infty} \log \left(\frac{\mathcal{K}_{\frac{1}{2}-n} \left(\frac{x}{b} \right)}{\mathcal{K}_{\frac{1}{2}-n} \left(\frac{x+m}{b} \right)} \left(\frac{x}{x+m} \right)^{n-\frac{1}{2}} \right) \\ &= \frac{m}{b}. \end{aligned}$$

By the definition of b , we have $\frac{m}{b} = \epsilon$. Therefore

$$\epsilon = \lim_{x \rightarrow \infty} \log \left(\frac{p(x)}{p(x+k)} \right) \leq \max_{k \in [-m, m]} \log \left(\frac{p(x)}{p(x+k)} \right) \leq \epsilon'.$$

Thus, we conclude that $\epsilon' = \epsilon$. \square

(ϵ, δ) -DP Gaussian Mechanism Amplification

Denote the non-private hybrid mean estimator as

$$\hat{\mu}_H(w) := \frac{w}{cn} \sum_{i \in T} x_i + \frac{1-w}{(1-c)n} \sum_{i \in L} x_i$$

and the joint privacy noise (without the largest adversarial coalition) as

$$Y := wY_T + \frac{1-w}{(1-c)n} \sum_{i \in L \setminus A} Y_{L,i}.$$

We first compute the sensitivity

$$\Delta_2 \hat{\mu}_H(w) = \max \left\| \hat{\mu}_H(w) - \hat{\mu}'_H(w) \right\|_2,$$

where $\hat{\mu}_H(w)$ is the estimator on any dataset $D = T \cup L$ and $\hat{\mu}'_H(w)$ is the estimator on any neighboring

dataset $D' = T' \cup L'$ differing in the data of at most one user. If the data of one T user is changed, then $\max \|\hat{\mu}_H(w) - \hat{\mu}'_H(w)\|_2 \leq \frac{wm}{cn}$. If instead the data of one L user is changed, then $\max \|\hat{\mu}_H(w) - \hat{\mu}'_H(w)\|_2 \leq \frac{(1-w)m}{(1-c)n}$. Note that $\frac{wm}{cn} \leq \frac{(1-w)m}{(1-c)n}$ when $w \leq c$. Thus, we have

$$\Delta_2 \hat{\mu}_H(w) = \begin{cases} \frac{wm}{cn}, & w \leq c \\ \frac{(1-w)m}{(1-c)n}, & \text{otherwise} \end{cases}.$$

Next, let $Y_T \sim \mathcal{N}(0, s_T^2)$ and $Y_{L,i} \sim \mathcal{N}(0, s_{L,i}^2)$ such that s_T satisfies (ϵ, δ) -DP for the TCM group and $s_{L,i}$ satisfies (ϵ, δ) -DP for each LM user i . By the well-known properties of Gaussians, the weighted combination of Gaussians is also a Gaussian, as

$$Y \sim \mathcal{N}\left(0, s'^2 := w^2 s_T^2 + \left(\frac{1-w}{1-c}\right)^2 |L \setminus A| s_L^2\right).$$

Recall that the classic Gaussian mechanism [24] guarantees (ϵ, δ) -DP for a function f with sensitivity $\Delta_2 f$ by adding noise from $\mathcal{N}(0, s^2)$ such that $s^2 = 2 \log(1.25/\delta) \Delta_2 f^2 / \epsilon^2$. Applying this result to our problem with a fixed $\delta' = \delta$ and solving ϵ' , we have

$$\begin{aligned} \epsilon' &= \sqrt{\frac{2 \ln(1.25/\delta) \Delta_2 \hat{\mu}_H^2(w)}{s'^2}} \\ &= \frac{\sqrt{2 \ln(1.25/\delta) m}}{ns'} \cdot \begin{cases} \frac{w}{c}, & w \leq c \\ \frac{1-w}{1-c}, & \text{otherwise} \end{cases}. \quad \square \end{aligned}$$

E Hybrid K-means Pseudocode

Algorithm 1: Hybrid-DP K -means

Input: TCM users T , LM users L , data range m , data dimension d , num. clusters K , num. iterations τ

- 1 Initialize centers of clusters C_1, \dots, C_K
- 2 $b_T \leftarrow \frac{(md+1)\tau}{\epsilon}$ and $b_L \leftarrow \frac{md(\tau+1)}{\epsilon}$
- 3 Each $i \in T$ reports $\tilde{x}_i \leftarrow x_i$ to the curator
- 4 Each $i \in L$ reports $\tilde{x}_i \leftarrow x_i + Y_{L,i}$ to the curator, $Y_{L,i} \sim \text{Lap}^d(b_L)$
- 5 **for** $t \leftarrow 1 \dots \tau$ **do**
- 6 Assign each \tilde{x}_i from T to closest cluster non-privately
- 7 Assign each \tilde{x}_i from L to closest cluster with prob. $\frac{\exp(\epsilon/(\tau+1))-1}{K+\exp(\epsilon/(\tau+1))-1}$; to a uniformly random cluster otherwise
- 8 **for** $k \leftarrow 1 \dots K$ **do**
- 9 Count T users in cluster k with DP : $\tilde{N}_T \leftarrow |C_k \cap T| + Y_1$, $Y_1 \sim \text{Lap}(b_T)$
- 10 Compute mean of all T users' data in cluster k with DP: $\tilde{\mu}_T \leftarrow \frac{1}{\tilde{N}_T} (\sum_{i \in T} x_i + Y_2)$, $Y_2 \sim \text{Lap}^d(b_T)$
- 11 Count L users in cluster k : $\tilde{N}_L \leftarrow |C_k \cap L|$
- 12 Compute mean of all L users' data in cluster k : $\tilde{\mu}_L \leftarrow \frac{1}{\tilde{N}_L} \sum_{i \in L} \tilde{x}_i$
- 13 $c \leftarrow \frac{\tilde{N}_T}{\tilde{N}_T + \tilde{N}_L}$, $s_T^2 \leftarrow 2b_T^2$, and $s_L^2 \leftarrow 2b_L^2$
- 14 Compute w_{PWH} as defined in Def. 5.3
- 15 Update center of C_k to hybrid mean of all data: $w_{PWH} \tilde{\mu}_T + (1 - w_{PWH}) \tilde{\mu}_L$
- 16 **return** centers of C_1, \dots, C_K
