High-Probability Bounds For Heterogeneous Local Differential Privacy

Maryam Aliakbarpour
Rice University
Ken Kennedy Institute

Alireza Fallah
Rice University
Ken Kennedy Institute

Swaha Roy
Rice University

Ria Stevens
Rice University

Abstract

We study statistical estimation under local differential privacy (LDP) when users may hold heterogeneous privacy levels and accuracy must be guaranteed with high probability. Departing from the common in-expectation analyses, and for one-dimensional and multi-dimensional mean estimation problems, we develop finite sample upper bounds in ℓ_2-norm that hold with probability at least $1-\beta$. We complement these results with matching minimax lower bounds, establishing the optimality (up to constants) of our guarantees in the heterogeneous LDP regime. We further study distribution learning in ℓ_∞-distance, designing an algorithm with high-probability guarantees under heterogeneous privacy demands. Our techniques offer principled guidance for designing mechanisms in settings with user-specific privacy levels.

1 INTRODUCTION

The unprecedented growth of data collection has made protecting user privacy a central challenge. Local Differential Privacy (LDP) offers a compelling solution, enabling the analysis of population-level statistics without exposing any individual's raw data—even to the aggregator (Dwork et al., 2006; Kasiviswanathan et al., 2011). In this model, each user perturbs their own data before transmission, ensuring that only randomized reports reach the curator. This paradigm has moved well beyond theory, with large-scale deployments at Google (Erlingsson et al., 2014), Microsoft (Ding et al., 2017), and Apple (Differential Privacy Team, Apple, 2017; Thakurta et al., 2017).

Most research on LDP, however, rests on two simplifying assumptions: that all users share a uniform privacy guarantee (ε) and that error bounds only need to hold in expectation or just with a constant probability. In this work, we move beyond this idealized model to study two crucial and more realistic variants of LDP.

The first variant is the *heterogeneous privacy setting*, where each of the n users may have their own distinct privacy parameter ε_i. This framework acknowledges that in practice, populations are diverse: some users may demand stronger protections, while others are willing to trade privacy for utility. While tuning local perturbation protocols to different ε_i is straightforward, accurately aggregating these heterogeneous reports into global estimates poses significant algorithmic challenges.

The second, and complementary, variant concerns the need for *high-probability results*—that is, error guarantees that hold with probability at least $1-\beta$. The conventional approach to achieving such results is to apply a standard amplification technique, like the median-of-means method, to an algorithm with in-expectation guarantees (e.g., Proposition 9 in (Hsu and Sabato, 2016)). However, this off-the-shelf approach suffers from two critical drawbacks:

1. *Suboptimal Dependence on Confidence:* These generic methods yield a poor dependence on the confidence parameter β. They typically multiply the overall error by a factor polynomial in $\log(1/\beta)$, whereas an ideal method's error would increase only by an additive term involving $\log(1/\beta)$.

2. *Breakdown of Repetition in Heterogeneous Settings:* These amplification techniques rely on a divide-and-conquer strategy, partitioning the population into subgroups to generate independent estimates. This strategy fails in the heterogeneous setting because it is non-trivial to ensure each subgroup has

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300.

a statistically similar composition of privacy budgets. For example, if a few users have very large ε_i values, they cannot be distributed evenly. The estimates from the subgroups containing them will be fundamentally different from the others, thereby breaking the statistical symmetry required for the median-of-means approach to work.

While some prior work has studied heterogeneous LDP (Fallah et al., 2024; Chaudhuri and Courtade, 2023; Chaudhuri et al., 2025), its error bounds only hold in expectation. For the very reasons just described—particularly the breakdown of repetition-based approaches—these guarantees cannot be readily converted into high-probability results, leaving the challenge of achieving robust, high-probability guarantees in this setting unaddressed.

In this work, we initiate the first systematic study that tackles these challenges simultaneously. We develop methods for heterogeneous LDP that provide tight high-probability error bounds without resorting to computationally expensive or statistically loose amplification techniques. Specifically, for both *single- and multi-dimensional mean estimation*, we establish sharp characterizations of the trade-off between heterogeneous privacy budgets and estimation error. We design computationally efficient algorithms and prove matching lower bounds, demonstrating that our results are optimal up to constant factors. Moreover, our bounds achieve the desired additive logarithmic dependence on the confidence parameter β. Notably, for the high-dimensional LDP mean estimation problem, to our knowledge, no prior analysis has established optimal dependence on all parameters—even in the standard uniform-privacy setting.

Furthermore, as a key application of our results, we use our single-dimension mean estimator to address the problem of *distribution learning* in ℓ_∞-distance. We apply our techniques to the framework of Bassily and Smith (2015), which relies on multiple mean estimation subroutines that must all be correct with high probability. Our high-probability estimation techniques naturally extend to this setting, yielding a new upper bound for distribution learning under heterogeneous local privacy.

1.1 Problem Formulation

We begin by defining local differential privacy:

Definition 1 (Pure Local Differential Privacy (LDP) (Kasiviswanathan et al., 2011)). *An algorithm $\mathcal{C} : \mathcal{X} \to \mathcal{Y}$ is said to be ε-locally differentially private if for any measurable set $\mathcal{W} \subseteq \mathcal{Y}$ and every pair of points $X, X' \in \mathcal{X}$,*

$$\mathbf{Pr}[\mathcal{C}(X) \in \mathcal{W}] \leq e^{\varepsilon}\, \mathbf{Pr}[\mathcal{C}(X') \in \mathcal{W}]. \tag{1}$$

Consider a setting with n users, each holding a data point X_i drawn i.i.d. from an unknown distribution P with parameter $\theta = \theta(P)$. Each user specifies a privacy parameter ε_i, indicating the level of protection they require. We denote the vector of these privacy parameters by $\boldsymbol{\varepsilon} \coloneqq \{\varepsilon_i\}_{i=1}^n$. They transmit a perturbed report Y_i, which is an ε_i-locally differentially private version of X_i. The data curator receives $\{Y_i\}_{i=1}^n$ and computes an estimator $\hat{\theta}$ of θ.

Our goal is to design a protocol for estimating θ that respects each user's privacy constraint. Formally, we construct n privacy channels $\{\mathcal{C}_i\}_{i=1}^n$, where each $\mathcal{C}_i : \mathcal{X}_i \to \mathcal{Y}_i$ satisfies ε_i-LDP with respect to user i's data.

We measure the quality of our estimator by finding an upper bound on the error of the estimation that holds with probability $1-\beta$. In particular, for any given $\beta \in (0,1)$, we are looking for the smallest $t \geq 0$ for which:

$$\mathbf{Pr}\Big[\mathrm{err}\left(\hat{\theta}(\boldsymbol{\mathcal{C}}(\boldsymbol{X})) - \theta\right) \leq t\Big] \geq 1-\beta, \tag{2}$$

where $\boldsymbol{\mathcal{C}}(\boldsymbol{X}) \coloneqq (\mathcal{C}_1(X_1), \ldots, \mathcal{C}_n(X_n))$. We examine three specific variations of this problem. In each problem, the heterogeneous set-up remains the same, but the nature of the underlying distribution or the parameter may be different.

Single-Dimensional Mean Estimation: Each user $i \in [n]$ holds a data point X_i drawn from a distribution P with mean θ over a bounded single-dimensional domain (either $[-1,+1]$ or $\{-1,+1\}$). The server estimates the mean of P by aggregating each user's privatized data and outputting $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$.

Multi-Dimensional Mean Estimation: Each user $i \in [n]$ holds a data point in the ℓ_2 (Euclidean) ball of radius r, $X_i \in \mathbb{B}^d(r)$, drawn from a distribution P with mean θ. The server estimates the mean of P by aggregating each user's data and outputting $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$.

Distribution Learning: Each user $i \in [n]$ holds a datapoint $x_i \sim P$, where P is a distribution over $[d]$. The server constructs an estimator $\hat{P} : [d] \to \mathbb{R}$ such that $\hat{P}(v)$ approximates $P(v)$ for all $v \in [d]$.

1.2 Our Contributions

Single-Dimensional Mean Estimation. We study mean estimation for single-dimensional

bounded random variables, $X_i \in [-1, +1]$. Our core contribution in this setting is an *optimized weighted average* of Y_i's, where the weights are proportional to the square of the users' privacy budgets, ε_i^2. This choice improves accuracy by down-weighting noisier contributions from users with stronger privacy requirements (smaller ε_i), whose privatized data points are more distorted. Our approach leads to the following high-probability error bound.

Theorem 1 (Informal version of Theorem 2.1 and Theorem 2.2)**.** *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, where $\varepsilon_i \le 1$ for all i and $\beta \in (0,1)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, X is bounded almost surely. Then, for all $P \in \mathcal{P}$, there exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ such that with probability at least $1-\beta$,*

$$\left|\hat{\theta}\left(X_{1:n}\right) - \theta\left(P\right)\right|^2 \le \mathcal{O}\left(\min\left(\frac{\log\left(1/\beta\right)}{\sum_{i=1}^n \varepsilon_i^2}, 1\right)\right). \quad (3)$$

We show this bound is tight in Theorem 2.3 using a high-probability instance of Le Cam's Lemma (Ma et al., 2024). To establish the upper bound, we provide two distinct algorithms. For general bounded distributions, Theorem 2.1 provides a formal guarantee for an estimator (Algorithm 1) based on the *Laplace mechanism*. As this mechanism requires communicating continuous values, which can be inefficient, we also present a communication-optimal method for binary data. In this setting, Theorem 2.2 proves the same tight bound is achievable with an estimator (Algorithm 2) based on *Randomized Response*.

Multi-Dimensional Mean Estimation. Our second main contribution is to extend our analysis to the *multi-dimensional* setting, where we provide the first tight, high-probability bounds for mean estimation over the Euclidean ball under heterogeneous Local Differential Privacy (LDP).

Theorem 2 (Informal version of Theorem 3.1)**.** *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, where $\varepsilon_i \le 1$ for all i and $\beta \in (0,1)$. For all $P \in \mathcal{P}_{2,r}$ such that for any $X \sim P$, $X \in \mathbb{B}^d(r)$ almost surely, there exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ such that with probability at least $1-\beta$,*

$$\left\|\hat{\theta}\left(X_{1:n}\right) - \theta\left(P\right)\right\|_2^2 \le \mathcal{O}\left(\frac{r^2\left(d + \log\left(1/\beta\right)\right)}{\sum_{i=1}^n \varepsilon_i^2}\right). \quad (4)$$

We present an efficient algorithm (Algorithm 3) that, similar to our one-dimensional estimator, computes an *optimized weighted average* of privatized versions user data. Each user privatizes their high-dimensional data point using a mechanism of Duchi et al. (2013), which projects data points onto one of two hemispheres based on a randomized response selection. Again, our algorithm weights each received signal proportionally to ε_i^2 to construct the final estimate.

The key novelty of our work lies in the high-probability nature of our bound. While bounds on the *expected* error were known, converting them to high-probability guarantees is not straightforward, and standard techniques yield suboptimal error rates. Our main technical contribution is a refined analysis that overcomes this challenge. We exploit the concentration of measure phenomenon that arises due to the privacy mechanism taking the form of a mixture of uniform distributions over hemispheres. This allows us to obtain an upper bound with an optimal *additive* dependence on the dimension and the confidence parameter (i.e., $d + \log(1/\beta)$), a significant improvement over the suboptimal multiplicative dependence ($d \cdot \log(1/\beta)$) produced by standard approaches that do not take into account the geometry of the private mechanism.

Finally, in Theorem 3.2, we prove that our algorithm is optimal by establishing a *matching lower bound*. This lower bound relies on minimax quantiles—high-probability analogues of minimax risk—and is constructed via a combination of Assouad's Lemma and a high-probability version of Le Cam's Lemma presented in Ma et al. (2024). This result solidifies our upper bound and provides a complete characterization of the high-probability error for this fundamental high-dimensional estimation problem.

Distribution Learning. Our final contribution addresses distribution learning under heterogeneous local privacy. We adapt the projection-based frequency estimation algorithm of Bassily and Smith (2015), originally designed for the homogeneous setting, to distribution learning in the heterogeneous case. At a high level, our algorithm applies a Johnson–Lindenstrauss (JL) transform to compress user data into a lower-dimensional representation, privatizes a randomly selected coordinate via randomized response, and aggregates the resulting reports with carefully chosen weights to account for heterogeneous privacy levels. This yields a communication- and computation-efficient oracle to estimate the probabilities of the domain elements that enables distribution learning with ℓ_∞ error guarantees.

Theorem 3 (Informal version of Theorem 4.1)**.** *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, where $\varepsilon_i \le 1$ for all i. Let $\mathcal{P}$ be the family of distributions P over $[d]$. Then, for all $P \in \mathcal{P}$, there exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{P}$ such that*

$$\left\|\hat{P}\left(x_{1:n}\right) - P\right\|_\infty \le \mathcal{O}\left(\sqrt{\frac{\log\left(d\right)}{\sum_{i=1}^n \varepsilon_i^2}}\right). \quad (5)$$

This theorem shows that our method achieves optimal dependence on the mixed privacy budgets, matching the rates known for the homogeneous case.

1.3 Minimax Quantiles

Standard techniques for lower bounding the expected error of an algorithm fail to capture the behaviour of its tail and the dependence on β. Therefore, to establish our lower bounds, we must rely on minimax quantiles (Ma et al., 2024).

Definition 2 (LDP-Minimax Quantiles)**.** *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0,1)$. Let $\mathcal{P}$ be a family of distributions, where $P_\theta \in \mathcal{P}$ is parametrized by $\theta \in \Theta$. Let $\mathcal{Q}_{\boldsymbol{\varepsilon}}$ be the set of all conditional distributions $Q : \mathcal{X}^{\otimes n} \to \mathcal{Y}^{\otimes n}$ guaranteeing $\boldsymbol{\varepsilon}$-local differential privacy. Let $\hat{\Theta}$ be the set of all measurable functions $\hat{\theta} : \mathcal{Y}^{\otimes n} \to \Theta$. Then, we define the minimax quantile, $\mathcal{M}(\beta, \mathcal{P}, \boldsymbol{\varepsilon})$ as:*

$$\mathcal{M}(\beta, \mathcal{P}, \boldsymbol{\varepsilon}) \coloneqq \inf_{Q \in \mathcal{Q}_{\boldsymbol{\varepsilon}}} \inf_{\hat{\theta} \in \hat{\Theta}} \sup_{P \in \mathcal{P}} \inf \left\{ t \in [0, \infty) : \right.$$
$$\left. \mathbf{Pr}_{P,Q}\left[\left\| \hat{\theta}(Y_{1:n}) - \theta(P) \right\|_2^2 \leq t \right] \geq 1 - \beta \right\}. \quad (6)$$

By lower bounding the minimax quantile for a given problem by t, we are saying that no algorithm can achieve error better than t with probability at least $1 - \beta$ on that problem and over all distributions in family $\mathcal{P}$.

Ma et al. (2024) also introduce methods that may be used to lower bound minimax quantiles. Specifically, they introduce a high-probability analogue of Le Cam's Lemma and give results that allow us to translate from expectation lower bounds to lower bounds on minimax quantiles. Our results use these techniques, among others.

1.4 Notation

We denote by $[n]$ the set $\{1, \ldots, n\}$. We denote by $\mathbb{S}^{d-1}(r)$ and $\mathbb{B}^d(r)$ the Euclidean sphere and ball respectively of radius r in to $\mathbb{R}^d$. We use the notation $\theta(P)$ to indicate that the distribution P is parametrized by θ. In the context of our work, this means that θ is the expected value of P.

1.5 Related Works

Differential privacy, introduced by (Dwork et al., 2006), has been extensively studied across both the central and local model (Hsu et al., 2012; Bassily and Smith, 2015; Bassily et al., 2017; Acharya et al., 2019; Asoodeh et al., 2021; Canonne and Gentle, 2025). We refer the reader to detailed surveys for more information on the breadth of applications (Cormode et al., 2018; Wang et al., 2020; Yang et al., 2024).

We are primarily interested in the heterogeneous setting where users have an individual privacy preference. Alaggan et al. (2017) and Jorgensen et al. (2015) independently introduced this notion of heterogeneous differential privacy, also known as Personalized Differential Privacy in the central model. Chen et al. (2016) first extended this notion to the local model. Canonne and Sun (2024) studied closeness testing under the heterogeneous setting in the local and shuffle models, although in their work there are only two possible privacy parameters.

Recent work has focused on mean estimation under heterogeneous differential privacy. Fallah et al. (2022) studied this problem under the Rényi DP setting in both the local and central models, with results holding in expectation. Chaudhuri and Courtade (2023) and Chaudhuri et al. (2025) proposed algorithms for single-dimensional mean estimation under heterogeneous *central* differential privacy. They further presented the "saturation phenomenon" which demonstrates that relaxing the privacy requirements for some users while keeping the privacy parameter of the others fixed beyond a critical point does not improve accuracy of a heterogeneous mean estimator. Fallah et al. (2024) provided optimal mean estimators for heterogeneous privacy with matching minimax lower bounds in both the central and local model. Their results apply to the single-dimensional setting and hold in expectation. In contrast, we propose high probability bounds for both the single-dimensional and multi-dimensional settings. Chaudhuri and Courtade (2025) also studied the heterogeneous differential privacy in the central model in the case that the users' data is correlated with their privacy parameters.

In this work, we also study the problem of distribution learning, where we focus on estimating the true proportion of an underlying population. This problem is similar to count estimation, frequency estimation and the identification of "heavy hitters". Literature in this area focuses on optimizing various parameters including the worst-case error, time complexity for each user, time complexity at the server, and communication complexity. Chen et al. (2016) provided the first heterogeneous LDP algorithm for count estimation over a spatial domain. Our work differs from this result both in terms of setting and algorithm. While their aggregation technique is tailored to the hierarchical structure of spatial data, we develop a protocol well-suited to data from an arbitrary distribution.

Bassily and Smith (2015) presented the first LDP heavy hitters algorithm with optimal worst-case er-

ror. Bassily et al. (2017) then proposed an algorithm which significantly improved both server and user runtime. Both of these results rely on shared randomness which is referred to as the "public-coin" setting. Acharya et al. (2019) developed a "private-coin" (i.e., no reliance on shared randomness) construction that achieved sample order optimality, logarithmic communication, and nearly linear runtime. Recently, Canonne and Gentle (2025) examined the medium- or low-privacy (large ε) regime departing from the commonly studied high-privacy (small ε) regime and presented near-tight bounds on error for distribution learning.

2 MEAN ESTIMATION OF A SINGLE DIMENSION

In this section, we study the heterogeneous LDP mean estimation problem for bounded distributions of a single dimension. We assume that each of n users holds a data point X_i drawn from a distribution P over a bounded domain $[-1, 1]$. Each user passes their data through an ε_i-LDP channel before sending it to the server. The server then estimates the mean of P, θ, by aggregating each user's data. We find that different choices of ε_i across users would result in suboptimal tail bounds on the error if the server were to output an unweighted mean of the privatized data points. To negate this effect, we propose that the server computes an weighted average of the users' data, with weights chosen according to each user's privacy parameter.

We give matching upper and lower bounds for the heterogeneous-LDP mean estimation problem. We provide two algorithms achieving our upper bounds: the first, Algorithm 1, applies to all distributions over $[-1, 1]$; and the second, Algorithm 2, specifically applies to binary distributions and achieves a low communication cost. We construct a hard instance and apply a high-probability instance of Le Cam's Lemma (Ma et al., 2024) to prove our lower bound.

2.1 Upper Bounds for Bounded Distributions

Algorithm 1 uses a Laplace mechanism to guarantee the privacy of each user and outputs a weighted sum of each user's privatized data. Using standard concentration bounds, we give tail bounds of its error. The proof is deferred to Section B.1.

Theorem 2.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \le 1$ for all i. Let $\beta \in (0, 1)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, $X \in [-1, 1]$ almost surely. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ and a universal constant c such that, for all $P \in \mathcal{P}$,*

$$\mathbf{Pr}_{X_{1:n}\sim P^{\otimes n}}\left[\left|\hat{\theta}\left(X_{1:n}\right) - \theta\left(P\right)\right|^2 \le \min\left(c\,\frac{\log\left(1/\beta\right)}{\sum_{i=1}^n \varepsilon_i^2}, 1\right)\right] \ge 1 - \beta. \quad (7)$$

Algorithm 1 achieves this bound.

Algorithm 1 Laplace Mechanism for Heterogeneous Local Differential Privacy

Require: Each user i has private data $X_i \in [-1, 1]$ and privacy parameter $\varepsilon_i > 0$.
Ensure: Estimate $\hat{\theta}$.
1: **for** each user $i = 1, \ldots, n$ **do**
2: Draw $Z_i \sim \mathbf{Lap}(2/\varepsilon_i)$.
3: $Y_i \leftarrow X_i + Z_i$.
4: Send Y_i to the server.
5: **end for**
6: Server sets $w_i \leftarrow \left(1 + \frac{1}{\varepsilon_i^2}\right)^{-1} / \sum_{j=1}^n \left(1 + \frac{1}{\varepsilon_j^2}\right)^{-1}$ for all i.
7: Server outputs $\hat{\theta} \leftarrow \sum_{i=1}^n w_i Y_i$.

2.2 Upper Bounds for Discrete Distributions

The guarantees of Algorithm 1 apply to any bounded single-dimensional distribution, including binary distributions. However, this mechanism requires that each user communicates a continuous value, Y_i, with the server, regardless of whether that user's data was originally continuous. When each user holds only one bit of data, it is desirable to design a more communication-efficient mechanism, wherein each user only communicates one bit with the server.

To address this challenge, we propose a heterogeneous LDP mean estimation mechanism, based on the randomized response mechanism. Specifically, assume each $X_i \sim P$, where $\mathbb{E}[P] = \theta \in [-1, 1]$ and $X_i \in \{-1, +1\}$. We construct a mean estimation mechanism that requires only one bit of communication per user, while achieving the same high-probability error guarantees as Algorithm 1.

As in the previous algorithm, the proposed mechanism involves user i sending the server an ε_i-private copy of their data. The server then outputs a weighted sum of their data, with weights determined according to each user's privacy level. This mechanism is presented in Algorithm 2. In Theorem 2.2, we establish tail bounds on its error. This theorem is proved in Section B.2.

Theorem 2.2. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \le 1$ for all i. Let $\beta \in (0, 1)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, $X \in \{-1, 1\}$ almost surely. There exists an $\boldsymbol{\varepsilon}$-locally differentially private*

Algorithm 2 Randomized Response Mechanism for Heterogeneous Local Differential Privacy

Require: Each user i has private data $X_i \in \{-1, 1\}$ and privacy parameter $\varepsilon_i > 0$.
Ensure: Estimate $\hat{\theta}$
1: **for** each user $i = 1, \ldots, n$ **do**
2: $Y_i \leftarrow \begin{cases} X_i & w.p. \ \frac{e^{\varepsilon_i}}{e^{\varepsilon_i}+1} \\ -X_i & w.p. \ \frac{1}{e^{\varepsilon_i}+1} \end{cases}$.
3: Send Y_i to the server.
4: **end for**
5: Server sets $c_i \leftarrow \frac{e^{\varepsilon_i}+1}{e^{\varepsilon_i}-1}$ for all i.
6: Server sets $w_i \leftarrow (1/c_i^2)/\sum_{j=1}^n (1/c_j^2)$ for all i.
7: Server outputs $\hat{\theta} \leftarrow \sum_{i=1}^n w_i c_i Y_i$.

estimator $\hat{\theta}$ and a universal constant c such that, for all $P \in \mathcal{P}$,

$$\mathbf{Pr}_{X_{1:n}\sim P^{\otimes n}}\left[\left|\hat{\theta}\left(X_{1:n}\right)-\theta\left(P\right)\right|^2 \leq \min\left(c\,\frac{\log\left(1/\beta\right)}{\sum_{i=1}^n \varepsilon_i^2}, 1\right)\right] \geq 1-\beta. \quad (8)$$

Algorithm 2 achieves this bound.

2.3 Lower Bounds

In Theorem 2.3, we establish a lower bound on the high probability error of one-dimensional mean estimation under heterogeneous LDP. Up to a constant factor, the upper bounds presented in Theorem 2.1 and Theorem 2.2 for the heterogeneous Laplace mechanism and randomized response, respectively, match this lower bound, thereby characterizing the tightness of these results.

Theorem 2.3. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0, 1/2)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, $X \in \{-1, 1\}$ almost surely. For the $\boldsymbol{\varepsilon}$-locally differentially private mean estimation problem over a single dimension, there exists an absolute constant c such that the minimax quantile is lower bounded as*

$$\min\left(c\,\frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2}, 1\right) \leq \mathcal{M}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right). \quad (9)$$

where $c > 0$ is a universal constant.

The proof of Theorem 2.3 is given in Section B.3. Briefly, our proof constructs a hard instance consisting of a pair of distributions, P_1 and P_2, whose means differ by a fixed gap. We apply a strong data processing inequality of Duchi et al. (2013) to show that passing these distributions through homogeneous privacy channels reduces their KL divergence, making them harder to distinguish. We then invoke a high probability version of Le Cam's Lemma, due to Ma et al. (2024), which implies that, because the privatized distributions are nearly indistinguishable, no algorithm can estimate their means accurately with high probability.

Algorithm 3 Optimal Multi-Dimensional Mean Estimation Mechanism under Heterogeneous Local Differential Privacy

Require: Each user i has private data $X_i \in \mathbb{B}^d(r)$, $X_i \sim P$, and privacy parameter $\varepsilon_i > 0$.
Ensure: Estimate $\hat{\theta}$ of $\mathbb{E}[P]$.
1: **for** each user $i = 1, \ldots, n$ **do**
2: $Y_i \leftarrow$ output of Strategy A of Duchi et al. (2013) (Algorithm 4) on X_i, ε_i.
3: Send Y_i to the server.
4: **end for**
5: Server sets $w_i \leftarrow \varepsilon_i^2/\sum_{j=1}^n \varepsilon_j^2$ for all i.
6: Server outputs $\hat{\theta} \leftarrow \sum_{i=1}^n w_i Y_i$.

3 MEAN ESTIMATION OF MULTIPLE DIMENSIONS

In this section, we study the heterogeneous LDP mean estimation problem for distributions over the Euclidean ball of radius r. We assume that each of n users holds a datapoint $X_i \in \mathbb{B}^d(r)$, drawn from a distribution P with mean θ. Each user passes their data through a locally private channel before sharing it with the untrusted server for aggregation. In Theorem 3.1, we give an upper bound for this problem, achieved by Algorithm 3. In Theorem 3.2, we prove a matching lower bound.

3.1 Upper Bounds

Within Algorithm 3, each user employs a locally differentially private mechanism proposed by Duchi et al. (2013) (Algorithm 4). On input $X_i \in \mathbb{B}^d(r)$, and given privacy parameter ε_i, this mechanism outputs $Y_i \in \mathbb{S}^{d-1}(B_i)$ for some $B_i = \mathcal{O}(rd/\varepsilon_i)$. In particular, the mechanism picks either the hemisphere $\left\{Y \in \mathbb{S}^{d-1}(B_i) \text{ s.t. } \langle Y, X_i\rangle > 0\right\}$ with some probability $p_i \geq 1/2$, or its complementary hemisphere with the complementary probability, and then samples uniformly from the chosen hemisphere. The choice of p_i ensures privacy, and together with the choice of B_i makes the output unbiased, i.e., $\mathbb{E}[Y_i] = X_i$. Lastly, as in the single-dimensional setting, the server outputs a weighted average of the users' projections, with the weight of user i being proportional to ε_i^2.

Algorithm 4 Locally-Differentially Private Randomizer of Duchi et al. (2013)

Require: Data point $X_i \in \mathbb{B}^d(r)$, privacy parameter $\varepsilon_i > 0$.

Ensure: Y_i, an ε_i-locally differentially private estimate of X_i.

$\tilde{X}_i \leftarrow \begin{cases} \frac{rX_i}{\|X_i\|_2} & w.p.\ \frac{1}{2} + \frac{\|X_i\|_2}{2r} \\ -\frac{rX_i}{\|X_i\|_2} & w.p.\ \frac{1}{2} - \frac{\|X_i\|_2}{2r} \end{cases}$.

$T \sim \textbf{Bernoulli}\left(\frac{e^{\varepsilon_i}}{e^{\varepsilon_i}+1}\right)$.

$c_i \leftarrow \frac{e^{\varepsilon_i}+1}{e^{\varepsilon_i}-1}$.

$B_i \leftarrow c_i r \frac{d\sqrt{\pi}\Gamma\left(\frac{d-1}{2}+1\right)}{\Gamma\left(\frac{d}{2}+1\right)}$.

if $T = 1$ **then**

  $Y_i \sim \textbf{Unif}\left(Y \in \mathbb{S}^{d-1}(B_i) \text{ s.t. } \left\langle Y, \tilde{X}_i \right\rangle > 0\right)$.

else if $T = 0$ **then**

  $Y_i \sim \textbf{Unif}\left(Y \in \mathbb{S}^{d-1}(B_i) \text{ s.t. } \left\langle Y, \tilde{X}_i \right\rangle \le 0\right)$.

end if

Output Y_i.

Theorem 3.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \le 1$ for all i. Let $\beta \in (0,1)$. Let $\mathcal{P}_{2,r}$ be the family of distributions P such that for any $X \sim P$, $X \in \mathbb{B}^d(r)$ almost surely. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ and a universal constant c such that, for all $P \in \mathcal{P}_{2,r}$,*

$$\mathbf{Pr}_{X_{1:n}\sim P^{\otimes n}}\left[\left\|\hat{\theta}(X_{1:n}) - \theta(P)\right\|_2^2 \le r^2 \min\left(c\frac{(d+\log(1/\beta))}{\sum_{i=1}^n \varepsilon_i^2}, 1\right)\right] \ge 1-\beta. \quad (10)$$

Algorithm 3 achieves this bound.

Proof Sketch (full proof in Section C.1). A natural attempt, similar to what we did in the single-dimensional case, would be to apply concentration inequalities for high-dimensional random variables. Given the boundedness of the output of the privacy channel, i.e., Y_i, two natural candidates are the Vector Bernstein inequality (Gross, 2011) and subgaussian bounds for bounded-support distributions. One can verify that both approaches lead to an upper bound with dependence on $d\log(1/\beta)$ (rather than what we ultimately show, which is $d + \log(1/\beta)$), which is suboptimal in the high-probability regime.

We resolve this issue by taking advantage of the structure of the ε_i-LDP randomizer of Duchi et al. (2013) (Algorithm 4)—specifically, that it draws Y_i from a mixture of uniform distributions over hemispheres. As we will discuss, uniform distributions over spheres exhibit the *concentration of measure* phenomenon, which allows us to derive tighter concentration inequalities.

Our point of departure in the proof is to break the problem of deriving high-probability bounds into two separate tail bounds. The first concerns the sampling (non-private) error of the algorithm, i.e., the difference between the weighted average of the X_i's and θ. The second concerns the error due to privacy, i.e., the difference between the the weighted average of the X_i's and the weighted average of the Y_i's, for any realization of the X_i's.

To bound the sampling error, we use the *norm-subgaussian* definition from Jin et al. (2019) (also stated in the appendix for completeness). In particular, the norm of each vector $X_i - \theta$ is bounded by $2r$, which implies that $X_i - \theta$ is a norm-subgaussian random vector with parameter $2r$. We can then apply the Hoeffding-like concentration inequality developed by Jin et al. (2019), which allows us to establish a bound of $r^2\log(d/\beta)/\sum_i \varepsilon_i^2$.

Bounding the tail of the privacy error (conditioned on realization of X_i's) is more nuanced. In particular, we show that, for any fixed i, $Y_i - X_i$ is subgaussian, with a subgaussian norm *independent of* d, which in turn leads to the desired high-probability bounds by applying norm concentration results for subgaussian vectors (e.g., see (Wainwright, 2019; Liu et al., 2025)). Recall that a random vector Z is subgaussian if $\langle Z, \ell\rangle$ is subgaussian for all ℓ, and the subgaussian norm of Z is defined as the supremum of the subgaussian norms of $\langle Z, \ell\rangle$ over all ℓ.

We establish this result using Lévy's lemma, a concentration-of-measure result on spheres that allows us to prove the sub-Gaussianity of Lipschitz functions of uniformly distributed random vectors on the sphere.

Lemma 3.1 (Levy's Lemma (Vershynin, 2018, Theorem 5.1.3))**.** *Let* **Unif** *be the uniform distribution over $\mathbb{S}^{d-1}(B)$. Let $f : \mathbb{S}^{d-1}(B) \to \mathbb{R}$ be an η-Lipschitz function. For all $t \ge 0$, there exists a universal constant c such that*

$$\mathbf{Pr}_{Y\sim\textbf{Unif}}[|f(Y) - \mathbb{E}[f(Y)]| \ge t] \le \exp\left(-c\frac{dt^2}{B^2\eta^2}\right).$$

This lemma allows us to establish subgaussian tail bounds on $f(Y) - \mathbb{E}[f(Y)]$ with $f(Y) = \langle \ell, Y\rangle$ when Y is drawn uniformly over $S = \mathbb{S}^{d-1}(B_i)$. However, in our setting, Y_i is not uniformly drawn over S. That said, we can interpret the distribution of Y_i as a mixture of a uniform distribution over S and a uniform distribution over S_1, where S_1 is the half-sphere of S defined by $\langle Y, X_i\rangle \ge 0$. Our proof proceeds in two steps:

first, we show that $f(Y)-\mathbb{E}[f(Y)]$ is subgaussian even when Y is drawn uniformly from the hemisphere S_1; second, combining the two results, we establish that the distribution of $\langle \ell, Y_i - X_i\rangle$ is subgaussian when Y_i is the output of the privacy channel of Algorithm 4.

To prove the first step, we define a function $g(Y)$ on the entire sphere S as follows: on S_1, let $g(Y) = \langle \ell, Y\rangle$; on S_2, first reflect Y into S_1 with respect to the plane separating S_1 and S_2 to obtain Y', and then let $g(Y) = \langle \ell, Y'\rangle$. Applying Lemma 3.1 to the function $g(\cdot)$ implies that the centered version of $\langle \ell, Y\rangle$ is subgaussian when Y is drawn uniformly from the hemisphere S_1.

For the second step, while we have shown that $f(Y) - \mathbb{E}[f(Y)]$ is subgaussian when Y is drawn uniformly from S or from S_1, the mean $\mathbb{E}[f(Y)]$ differs between these cases. This difference prevents us from immediately combining the tail bounds of both distributions to establish a tail bound on $f(Y) - \mathbb{E}[f(Y)]$ when $Y = Y_i$. Instead, we use the definition of subgaussianity based on moment-generating functions and show that the error introduced by these different means can be controlled by a constant factor. This, in turn, establishes the subgaussianity of $f(Y_i) - \mathbb{E}[f(Y_i)]$, completing the proof.

□

3.2 Lower Bounds

In Theorem 3.2, we give a lower bound on the high probability error of multi-dimensional mean estimation under heterogeneous LDP. This lower bound matches the upper bound presented in Theorem 3.1, demonstrating the optimality of Algorithm 3.

Theorem 3.2. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \le 1$ for all i. Let $\beta \in (0, 3/20)$. Let $r > 0$. For all $i \in [n]$, let $\varepsilon_i \in (0,1)$. For the mean estimation problem over the ℓ_2 ball with radius r, there exists an absolute constant c such that the minimax quantile is lower bounded as*

$$\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \ge c\, r^2 \min\left(\frac{\log(1/\beta)+d}{\sum_{i=1}^n \varepsilon_i^2}, \frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2} + \frac{1}{\sqrt{\sum_{i=1}^n \varepsilon_i^2}}, 1\right).$$

The proof of Theorem 3.2 is given in Section C.2. This proof involves constructing a hard instance, consisting of distributions parametrized by the vertices of the d-dimensional hypercube. We show that the distributions corresponding to any two adjacent vertices become nearly indistinguishable after they are passed through privacy channels. This allows us to apply Assouad's Lemma, therefore lower bounding the minimax risk of mean estimation over this class of distributions.

To connect this bound on minimax risk to minimax quantiles, we apply Theorem 6 (Ma et al., 2024).

4 DISTRIBUTION LEARNING

Algorithm 5 Distribution Learning Mechanism Under Local Heterogeneous Differential Privacy

Require: Each user i has private data $x_i \in [d]$ and privacy parameter $\varepsilon_i > 0$. Server has a global confidence parameter $\beta \in (0,1)$.
Ensure: Estimates $\hat{P}(v)$ for all $v \in [d]$.

$c_i \leftarrow \frac{e^{\varepsilon_i}+1}{e^{\varepsilon_i}-1}$ for all i.
$\gamma \leftarrow \frac{\sqrt{\log(2d/\beta)}}{\sum_{i=1}^n 1/c_i^2}$.
$m \leftarrow \frac{\log(d+1)\log(2/\beta)}{\gamma^2}$.
Server generates $\boldsymbol{\Phi} \sim \textbf{Unif}\left(\{-\frac{1}{\sqrt{m}}, \frac{1}{\sqrt{m}}\}^{m\times d}\right)$.
for each user $i = 1, \dots, n$ **do**
  $\mathbf{z}_i \leftarrow$ LOCAL-RANDOMIZER $(x_i, \varepsilon_i, \boldsymbol{\Phi}, c_i)$
  Send $\mathbf{z}_i$ to the server.
end for
Server sets $w_i \leftarrow (1/c_i^2)/\sum_{j=1}^n (1/c_j^2)$ for all i.
Server computes $\bar{\mathbf{z}} \leftarrow \sum_{i=1}^n w_i\, \mathbf{z}_i$.
Server outputs $\hat{P}(v) = \langle \bar{\mathbf{z}}, \boldsymbol{\Phi}\mathbf{e}_v\rangle$ for all $v \in [d]$.

The following subroutine is due to Bassily and Smith (2015).

procedure LOCAL-RANDOMIZER$(x_i, \varepsilon_i, \boldsymbol{\Phi}, c_i)$
  $\mathbf{y}_i \leftarrow \boldsymbol{\Phi}\mathbf{e}_{x_i}$.
  Sample $j \sim \textbf{Unif}([m])$.
  if $\mathbf{y}_i \neq \mathbf{0}$ **then**

$$\mathbf{z}_{ij} \leftarrow \begin{cases} c_i m\, \mathbf{y}_{ij} & w.p. \ \frac{e^{\varepsilon_i}}{e^{\varepsilon_i}+1} \\ -c_i m\, \mathbf{y}_{ij} & w.p. \ \frac{1}{e^{\varepsilon_i}+1} \end{cases}.$$

  else
    $\mathbf{z}_{ij} \sim \textbf{Unif}(\{-c_i\sqrt{m}, c_i\sqrt{m}\})$.
  end if
  Set $\mathbf{z}_{i\ell} = 0$ for all $\ell \neq j$.
  Send $\mathbf{z}_i$ to the server.
end procedure

In this section, we demonstrate an adaptation of the projection-based homogeneous locally private frequency estimation algorithm of Bassily and Smith (2015) for heterogeneous locally private distribution learning. Their algorithm produces a computation- and communication-efficient frequency oracle by employing a Johnson-Lindenstrauss (JL) transform to project the histogram of observed values to a lower-dimensional space.

More precisely, each user compresses their data point via a random $m \times d$ matrix $\Phi \sim \textbf{Unif}\{-1/\sqrt{m}, 1/\sqrt{m}\}$. Each user then selects

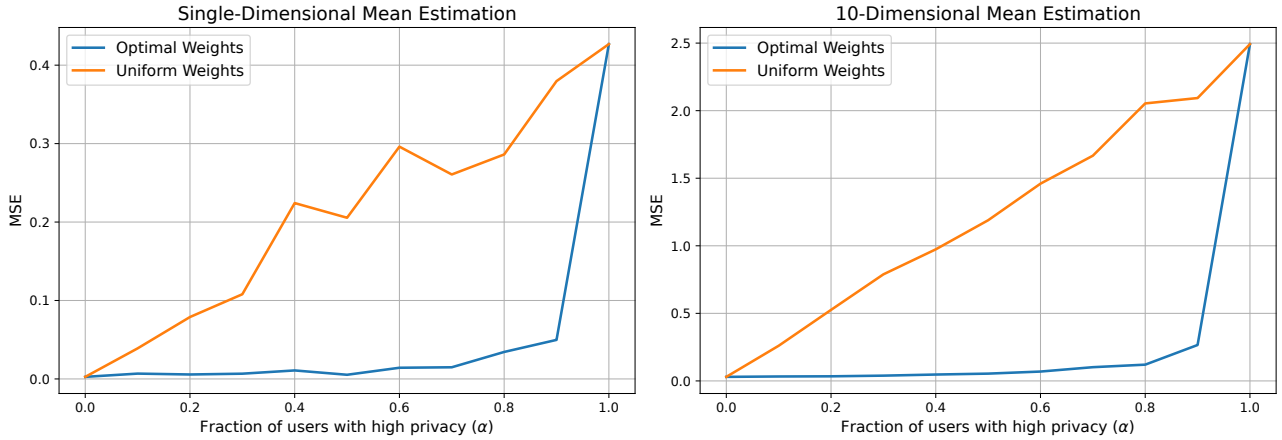

Figure 1: Comparison of mean-squared error for one-dimensional (left) and 10-dimensional (right) mean estimation on worst-case distributions under heterogeneous local differential privacy. Error is plotted as a function of α, the fraction of users with $\epsilon_i = 0.1$. The optimal weighting scheme proposed in this paper (blue) consistently outperforms the uniform weighting scheme (orange).

one bit of their compressed vector, runs randomized response to privatize this bit, and shares the resulting bit with the server, which aggregates all bits. The Johnson-Lindenstrauss Lemma (Lemma 4.1) ensures that the error due to this projection is not too large.

Lemma 4.1 (Johnson-Lindenstrauss Lemma)**.** *Let $0 < \lambda < 1$ and $d \in \mathbb{N}$. For any set $\mathcal{V}$ of t points in $\mathbb{R}^d$ and $m \geq \frac{8\log(t)}{\lambda^2}$, there exists a linear map $\Phi : \mathcal{R}^d \to \mathcal{R}^m$ such that for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, we have both:*

$$(1-\lambda)\left\|\mathbf{x}-\mathbf{y}\right\|_2^2 \leq \left\|\Phi(\mathbf{x}-\mathbf{y})\right\|_2^2 \leq (1+\lambda)\left\|\mathbf{x}-\mathbf{y}\right\|_2^2,$$
$$\text{and } \left|\langle \Phi\mathbf{x}, \Phi\mathbf{y}\rangle - \langle \mathbf{x}, \mathbf{y}\rangle\right| \leq O\left(\lambda\left(\left\|\mathbf{x}\right\|_2^2 + \left\|\mathbf{y}\right\|_2^2\right)\right).$$

Since we work in the heterogeneous LDP setting, we must adjust the server's aggregation step compared to the homogeneous case. We present the modified algorithm in Algorithm 5. In Theorem 4.1, we give guarantees on its ℓ_∞-error for distribution learning. This theorem is proved in Section D.1.

Theorem 4.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0,1)$. Let $\mathcal{P}$ be the family of distributions P over $[d]$. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{P}$ and a universal constant c such that, for all $P \in \mathcal{P}$,*

$$\mathbf{Pr}_{x_{1:n}\sim P^{\otimes n}}\left[\left\|\hat{P}\left(x_{1:n}\right) - P\right\|_\infty \leq \min\left(c\sqrt{\frac{\log\left(d/\beta\right)}{\sum_{i=1}^n \varepsilon_i^2}}, 1\right)\right] \geq 1-\beta. \quad (11)$$

Algorithm 5 achieves this bound.

We note that Algorithm 5 is not necessarily a proper learning algorithm, as there is no guarantee that the output $\hat{P}$ is a valid probability distribution. However, in Theorem 4.1, we show that it closely approximates the true distribution P in ℓ_∞ distance. To transform $\hat{P}$ into a probability distribution, one could project it onto the probability simplex without increasing the algorithm's error or sacrificing privacy.

We also remark that Algorithm 5 can be employed either to provide an estimate of the entire pmf of P or as an oracle for estimating the density of any given value, v, on demand. To perform the latter task, the algorithm does not need to execute Line 11 for all $v \in [d]$, therefore improving its computational efficiency.

5 EXPERIMENTS

We empirically demonstrate the utility of our optimal weighting technique for both one-dimensional and 10-dimensional mean estimation. We draw 1000 samples from the worst-case distributions described in our lower bounds, which are information-theoretically the most challenging. We consider a heterogeneous setting in which an α-fraction of the users demand strong privacy guarantees ($\varepsilon_i = 0.1$), while the remaining $(1-\alpha)$-fraction of the users have weaker privacy requirements ($\varepsilon_i = 1$). We compare the mean estimates produced by our algorithms with optimal weights to those obtained using uniform weights.

The results, given in Figure 1, show that our optimal weighting scheme achieves a lower error than uniform weights as more users demand higher privacy. Across all experiments, our algorithm effectively leverages the information contributed by users with lower privacy requirements ($\epsilon_i = 1$), leading to improved accuracy.

Acknowledgments

R.S. acknowledges partial support from the Ken Kennedy Institute Research Cluster Fund, the Ken Kennedy Institute Computational Science and Engineering Recruiting Fellowship, funded by the Energy HPC Conference and the Rice University Department of Computer Science, and the Ken Kennedy Institute 2025/26 Andrew Ladd Memorial Excellence in Computer Science Graduate Fellowship.

References

J. Acharya, Z. Sun, and H. Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1120–1129. PMLR, 2019.

M. Alaggan, S. Gambs, and A.-M. Kermarrec. Heterogeneous differential privacy. *Journal of Privacy and Confidentiality*, 7(2), Jan. 2017. doi: 10.29012/jpc.v7i2.652. URL https://journalprivacyconfidentiality.org/index.php/jpc/article/view/652.

S. Asoodeh, M. Aliakbarpour, and F. P. Calmon. Local differential privacy is equivalent to contraction of an f-divergence. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 545–550. IEEE, 2021.

R. Bassily and A. Smith. Local, private, efficient protocols for succinct histograms. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, page 127–135, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450335362. doi: 10.1145/2746539.2746632. URL https://doi.org/10.1145/2746539.2746632.

R. Bassily, K. Nissim, U. Stemmer, and A. Guha Thakurta. Practical locally private heavy hitters. *Advances in Neural Information Processing Systems*, 30, 2017.

C. L. Canonne and A. Gentle. Locally Private Histograms in All Privacy Regimes. In R. Meka, editor, *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, volume 325 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 25:1–25:24, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-361-4. doi: 10.4230/LIPIcs.ITCS.2025.25. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2025.25.

C. L. Canonne and Y. Sun. Private Distribution Testing with Heterogeneous Constraints: Your Epsilon Might Not Be Mine. In V. Guruswami, editor, *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*, volume 287 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 23:1–23:24, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-309-6. doi: 10.4230/LIPIcs.ITCS.2024.23. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2024.23.

T.-H. H. Chan, E. Shi, and D. Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3): 1–24, 2011.

S. Chaudhuri and T. A. Courtade. Mean estimation under heterogeneous privacy: Some privacy can be free. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 1639–1644. IEEE, 2023.

S. Chaudhuri and T. A. Courtade. Private Estimation When Data and Privacy Demands Are Correlated. In M. Bun, editor, *6th Symposium on Foundations of Responsible Computing (FORC 2025)*, volume 329 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 3:1–3:20, Dagstuhl, Germany, 2025. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-367-6. doi: 10.4230/LIPIcs.FORC.2025.3. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.FORC.2025.3.

S. Chaudhuri, K. Miagkov, and T. A. Courtade. Mean estimation under heterogeneous privacy demands. *IEEE Transactions on Information Theory*, 71(2): 1362–1375, 2025. doi: 10.1109/TIT.2024.3511498.

R. Chen, H. Li, S. P. Kasiviswanathan, and H. Jin. Private spatial data aggregation in the local setting. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 289–300. IEEE, 2016.

G. Cormode, S. Jha, T. Kulkarni, N. Li, D. Srivastava, and T. Wang. Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*, SIGMOD '18, page 1655–1658, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450347037. doi: 10.1145/3183713.3197390. URL https://doi.org/10.1145/3183713.3197390.

Differential Privacy Team, Apple. Learning with privacy at scale, 2017. URL https://machinelearning.apple.com/research/learning-with-privacy-at-scale.

B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th annual symposium on foundations of computer science*, pages 429–438. IEEE, 2013.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In S. Halevi and T. Rabin, editors, *Theory of Cryptography*, pages 265–284. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-32732-5.

Ú. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

A. Fallah, A. Makhdoumi, A. Ozdaglar, et al. Bridging central and local differential privacy in data acquisition mechanisms. *Advances in Neural Information Processing Systems*, 35:21628–21639, 2022.

A. Fallah, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Optimal and differentially private data acquisition: Central and local mechanisms. *Operations Research*, 72(3):1105–1123, 2024.

D. Gross. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory*, 57(3):1548–1566, 2011.

D. J. Hsu and S. Sabato. Loss minimization and parameter estimation with heavy tails. *J. Mach. Learn. Res.*, 17:18:1–18:40, 2016. URL https://jmlr.org/papers/v17/14-273.html.

J. Hsu, S. Khanna, and A. Roth. *Distributed Private Heavy Hitters*, page 461–472. Springer Berlin Heidelberg, 2012. ISBN 9783642315947. doi: 10.1007/978-3-642-31594-7_39. URL http://dx.doi.org/10.1007/978-3-642-31594-7_39.

C. Jin, P. Netrapalli, R. Ge, S. M. Kakade, and M. I. Jordan. A short note on concentration inequalities for random vectors with subgaussian norm. *arXiv preprint arXiv:1902.03736*, 2019.

Z. Jorgensen, T. Yu, and G. Cormode. Conservative or liberal? personalized differential privacy. In *2015 IEEE 31st International Conference on Data Engineering*, pages 1023–1034, 2015. doi: 10.1109/ICDE.2015.7113353.

S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Z. Liu, S. Power, and Y. Chen. A new proof of sub-gaussian norm concentration inequality. *arXiv preprint arXiv:2503.14347*, 2025.

T. Ma, K. A. Verchand, and R. J. Samworth. High-probability minimax lower bounds. *arXiv preprint arXiv:2406.13447*, 2024.

A. G. Thakurta, A. H. Vyrros, U. S. Vaishampayan, G. Kapoor, J. Freudiger, V. R. Sridhar, and D. Davidson. Learning new words, 2017. Publisher: United States Patent, Type: Patent.

R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

T. Wang, X. Zhang, J. Feng, and X. Yang. A comprehensive survey on local differential privacy toward data statistics and analysis. *Sensors*, 20(24):7030, Dec. 2020. ISSN 1424-8220. doi: 10.3390/s20247030. URL http://dx.doi.org/10.3390/s20247030.

M. Yang, T. Guo, T. Zhu, I. Tjuawinata, J. Zhao, and K.-Y. Lam. Local differential privacy and its applications: A comprehensive survey. *Computer Standards & Interfaces*, 89:103827, 2024.

B. Yu. Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam: research papers in probability and statistics*, pages 423–435. Springer, 1997.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Not Applicable]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:

(a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Not Applicable]

(b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Not Applicable]

(c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Not Applicable]

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

   (a) Citations of the creator If your work uses existing assets. [Not Applicable]

   (b) The license information of the assets, if applicable. [Not Applicable]

   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]

   (d) Information about consent from data providers/curators. [Not Applicable]

   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. [Not Applicable]

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A BACKGROUND RESULTS

In the following, we list useful facts, definitions, lemmas and theorems used throughout our proofs.

A.1 Subgaussian and Subexponential Random Variables

We give the definitions and key properties of subexponential and subgaussian random variables, as well as norm-subgaussian random vectors, used throughout our work. For further background on subexponential and subgaussian random vectors, see (Vershynin, 2018). For further background on norm-subgaussian random vectors, see Jin et al. (2019).

Definition 3 (Subexponential random variable)**.** *(Vershynin, 2018) Let $\nu, \alpha > 0$. A random variable with mean $\mathbb{E}[X] = \mu$ is (ν^2, α)-subexponential, denoted $X \in subE(\nu^2, \alpha)$ if*

$$\mathbb{E}\left[e^{\lambda(X-\mu)}\right] \leq e^{\frac{\nu^2\lambda^2}{2}} \qquad (\forall \lambda : |\lambda| \leq \tfrac{1}{\alpha})$$

Fact 1. *If $X \in subE(\nu^2, \alpha)$ then, for all $t > 0$,*

$$\mathbf{Pr}[|X - \mathbb{E}[X]| \geq t] \leq \exp\left(-\frac{1}{2}\min\left\{\frac{t^2}{\nu^2}, \frac{t}{\alpha}\right\}\right).$$

Definition 4 (Subgaussian random variable)**.** *(Vershynin, 2018) Let $\nu_1 > 0$. A random variable X is said to be ν_1-subgaussian if it satisfies:*

$$\mathbf{Pr}[|X| \geq t] \leq 2\exp\left(\frac{-t^2}{\nu_1^2}\right) \ \forall t \geq 0. \tag{12}$$

If $\mathbb{E}[X] = 0$, the property given in (12) is equivalent to the following for $\nu_2 = O(\nu_1)$:

$$\mathbb{E}[\exp(\lambda X)] \leq \exp\left(\nu_2^2\lambda^2\right) \ \forall \lambda \in \mathbb{R}. \tag{13}$$

We say that ν_1 (or ν_2) is the variance proxy *of X. We note that in (12), the constant 2 can be any absolute constant greater than 1. Similarly, in (13), a multiplicative constant other than 1 may be had. Changing these constants changes the variance proxy by a constant factor. (Remark 2.6.3 of Vershynin (2018)).*

Definition 5 (Subgaussian random vector)**.** *(Vershynin, 2018) Let $\nu > 0$. A random vector $X \in \mathbb{R}^d$ is said to be ν-subgaussian if $\langle \ell, X\rangle$ is ν-subgaussian for all $\ell \in \mathbb{S}^{d-1}$.*

Fact 2 (Norm of subgaussian random vectors)**.** *(Wainwright, 2019; Liu et al., 2025) For a subgaussian vector $X \in \mathbb{R}^d$ with variance proxy σ^2, there exist constants C_1, C_2 such that for all $\beta \in (0,1)$,*

$$\mathbf{Pr}\left[\|X\|_2 \leq \sigma\sqrt{C_1 d + C_2 \log(1/\beta)}\right] \geq 1 - \beta. \tag{14}$$

Definition 6. *Let $\sigma > 0$. A random vector $X \in \mathbb{R}^d$ is σ-norm-subgaussian if it satisfies, for all $t \in \mathbb{R}$,*

$$\mathbf{Pr}[\|X - \mathbb{E}[X]\|_2 \geq t] \leq 2\exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{15}$$

Fact 3. *Let $\sigma > 0$. Let $X \in \mathbb{R}^d$ be a random vector such that $\|X\| \leq \sigma$. Then, X is σ-norm-subgaussian.*

Fact 4. *(Jin et al., 2019) Let $\beta \in (0,1)$. Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be random vectors. Assume X_i is σ_i-norm-subgaussian for all $i \in [n]$. Then, there exists a constant c such that:*

$$\mathbf{Pr}\left[\left\|\sum_{i=1}^n X_i\right\|_2 \leq c\sqrt{\sum_{i=1}^n \sigma_i^2 \log(2d/\beta)}\right] \geq 1 - \beta \tag{16}$$

A.2 Lower Bound Techniques

Lemma 1 (Assouad's Lemma (Yu, 1997; Ma et al., 2024)). *Let $k \geq 1$. Let $\mathcal{V} = \{-1, 1\}^k$. For all $\nu, \nu' \in \mathcal{V}$, write $\nu \sim \nu'$ if ν and ν' differ in only one coordinate and $\nu \sim_j \nu'$ when that coordinate is j. For $\nu \in \mathcal{V}$, let $P_\nu \in \mathcal{P}$. Suppose there are k pseudo-distances $d_1, \ldots, d_k$ on Θ such that for any $\theta_1, \theta_2 \in \Theta$:*

$$d(\theta_1, \theta_2) = \sum_{j=1}^{k} d_j(\theta_1, \theta_2). \tag{17}$$

Then,

$$\inf_{\hat{\theta} \in \hat{\Theta}} \max_{\nu \in \mathcal{V}} \mathbb{E}_{X_{1:n} \sim P_\nu^{\otimes n}} \left[d\left(\hat{\theta}(X_{1:n}), \theta(P_\nu) \right) \right] \geq \frac{k}{2} \cdot \min_{j \in [k], \nu \sim_j \nu'} d_j(\theta(P_\nu), \theta(P_{\nu'})) \cdot \min_{\nu \sim \nu'} (1 - \mathrm{d_{TV}}(P_\nu, P_{\nu'})).$$

Theorem 4 (Pairwise upper bound on Kullback-Leibler divergences (Duchi et al., 2013, 2018)). *Let $\varepsilon \geq 0$. Consider any pair of distributions $\mu^{(1)}$ and $\mu^{(2)}$. Let $P^{(1)}$ and $P^{(2)}$ be the pair of distributions induced by passing samples drawn from $\mu^{(1)}$ and $\mu^{(2)}$ through a ε-LDP channel. Then we have*

$$\mathrm{KL}\left(P^{(1)} \,\middle\|\, P^{(2)}\right) + \mathrm{KL}\left(P^{(2)} \,\middle\|\, P^{(1)}\right) \leq \min\left\{4, e^{2\varepsilon}\right\} (e^\varepsilon - 1)^2 \, \mathrm{d}_{\mathrm{TV}}^2(P_1, P_2). \tag{18}$$

A.2.1 Minimax Quantiles

Ma et al. (2024) introduce the concept of minimax quantiles (Definition 2), which we use to construct our lower bounds. They further develop tools and techniques for working with minimax quantiles. Among these tools are lower minimax quantiles, defined in Definition 7. Lower minimax quantiles are particularly useful as they lower bound minimax quantiles, as stated in Theorem 6, and may be easier to work with. Corollary 1 and Theorem 6 give examples of techniques which may be used to lower bound lower minimax quantiles. Corollary 1 lower bounds this quantity whenever there exist two distributions in the family with sufficiently small KL divergence. Theorem 6 establishes a relationship between lower minimax quantiles and minimax risk.

Definition 7 (Lower minimax quantile (Ma et al., 2024)). *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i > 0$ for all i. Let $\beta \in (0, 1)$. Let $\mathcal{P}$ be a family of distributions, where $P_\theta \in \mathcal{P}$ is parametrized by $\theta \in \Theta$. Let $\mathcal{Q}_{\boldsymbol{\varepsilon}}$ be the set of all conditional distributions $Q : \mathcal{X}^{\otimes n} \to \mathcal{Z}^{\otimes n}$ guaranteeing $\boldsymbol{\varepsilon}$-local differential privacy. Let $\hat{\Theta}$ be the set of all measurable functions $\hat{\theta} : \mathcal{Z}^{\otimes n} \to \Theta$. Then, we define the lower minimax quantile, $\mathcal{M}_-(\beta, \mathcal{P}, \boldsymbol{\varepsilon})$ as:*

$$\mathcal{M}_-(\beta, \mathcal{P}, \boldsymbol{\varepsilon}) := \inf \left\{ t \in [0, \infty) \;:\; \inf_{Q \in \mathcal{Q}_{\boldsymbol{\varepsilon}}} \inf_{\hat{\theta} \in \hat{\Theta}} \sup_{P \in \mathcal{P}} \mathbf{Pr}_{P,Q}\left[\left\| \hat{\theta}(Z_{1:n}) - \theta(P) \right\|_2^2 \leq t \right] \geq 1 - \beta \right\}. \tag{19}$$

Theorem 5 (Theorem 4 of (Ma et al., 2024), adapted to our notation). *For all $\beta \in (0, 1]$, families of distributions $\mathcal{P}$ and privacy budgets $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$,*

$$\mathcal{M}_-(\beta, \mathcal{P}, \boldsymbol{\varepsilon}) \leq \mathcal{M}(\beta, \mathcal{P}, \boldsymbol{\varepsilon}). \tag{20}$$

Corollary 1 (Corollary 6 of Ma et al. (2024), adapted to our setting). *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \in (0, 1]$ for all i. Let $\beta \in (0, 1/2)$. Let $\boldsymbol{\mathcal{C}}$ be an $\boldsymbol{\varepsilon}$-privacy channel. Suppose that $P_1 \in \mathcal{P}$ and $P_2 \in \mathcal{P}$ satisfy $\mathrm{KL}(\boldsymbol{\mathcal{C}}(P_1) \,\|\, \boldsymbol{\mathcal{C}}(P_2)) < \log(1/(4\beta(1-\beta)))$. Then,*

$$\mathcal{M}_-(\beta, \mathcal{P}, \boldsymbol{\varepsilon}) \geq \left(\frac{\theta(P_1) - \theta(P_2)}{2} \right)^2. \tag{21}$$

Theorem 6 (Theorem 8 of Ma et al. (2024), adapted to our setting)**.** *Let $\Theta_0 \subseteq \Theta$ be non-empty. Let $\mathcal{P}_0 \subseteq \mathcal{P}$ such that for all $P \in \mathcal{P}_0$, $\theta(P) \in \Theta_0$. Define D^2 as*

$$D^2 := \sup_{P_1, P_2 \in \mathcal{P}_0} \|\theta(P_1) - \theta(P_2)\|_2^2. \tag{22}$$

Let $\Delta \in [0, \infty)$ be such that

$$\inf_{Q \in \mathcal{Q}_\varepsilon} \inf_{\hat{\theta} \in \hat{\Theta}} \sup_{P \in \mathcal{P}_0} \mathbb{E}_{P,Q}\left[\left\|\hat{\theta}(Z_{1:n}) - \theta(P)\right\|_2^2\right] \geq \Delta. \tag{23}$$

Then, if $D \neq 0$, for every $\varphi > 0$ and $\beta \in \left(0, \frac{\Delta - \varphi^2 D^2}{(1+\varphi)^2 D^2}\right)$, we have $\mathcal{M}_-(\beta, \mathcal{P}, \boldsymbol{\varepsilon}) \geq \varphi^2 D^2$.

B MISSING PROOFS FROM SECTION 2

B.1 Proof of Theorem 2.1

Theorem 2.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0,1)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, $X \in [-1,1]$ almost surely. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ and a universal constant c such that, for all $P \in \mathcal{P}$,*

$$\mathbf{Pr}_{X_{1:n} \sim P^{\otimes n}}\left[\left|\hat{\theta}(X_{1:n}) - \theta(P)\right|^2 \leq \min\left(c\,\frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2}, 1\right)\right] \geq 1 - \beta. \tag{7}$$

Algorithm 1 achieves this bound.

Proof. Consider Algorithm 1, wherein each user i adds $\mathbf{Lap}(2/\varepsilon_i)$ to their data, and the server generates an estimate, $\hat{\theta}$, by computing a weighted sum of the users' data, weighted by $w_i \propto \varepsilon_i^2$ for each user i.

We are hoping for a high-probability bound on the error of this algorithm, $\left|\hat{\theta} - \theta\right|^2$. By the triangle inequality and a union bound, we have:

$$\mathbf{Pr}\left[\left|\hat{\theta} - \theta\right|^2 > t^2\right] = \mathbf{Pr}\left[\left|\hat{\theta} - \theta\right| > t\right] \leq \mathbf{Pr}\left[\left|\sum_{i=1}^n w_i X_i - \theta\right| > \frac{t}{2}\right] + \mathbf{Pr}\left[\left|\sum_{i=1}^n w_i Z_i\right| > \frac{t}{2}\right]. \tag{24}$$

We begin by bounding the first term of (24). Observe that $\mathbb{E}\left[\sum_{i=1}^n w_i X_i\right] = \theta$, as $\sum_{i=1}^n w_i = 1$. Therefore, bounding the difference between $\sum_{i=1}^n w_i X_i$ and its expectation bounds the first term of (24). As $X_i \in [-1,1]$ almost surely, $w_i X_i \in [-w_i, w_i]$ and we can apply a Hoeffding bound. This results in:

$$\mathbf{Pr}\left[\left|\sum_{i=1}^n w_i X_i - \theta\right| > \frac{t}{2}\right] \leq 2\exp\left(-\frac{t^2}{8\sum_{i=1}^n w_i^2}\right). \tag{25}$$

To bound the second term of (24), define $\tilde{Z}_i := w_i Z_i$. Given that Z_i is a Laplace random variable with scale $2/\varepsilon_i$, $\tilde{Z}_i$ is Laplace with scale $2w_i/\varepsilon_i$. A standard argument from Chan et al. (2011) using the moment-generating functions of Laplace random variables shows that

$$\mathbb{E}\left[\exp\left(\lambda \tilde{Z}_i\right)\right] \leq \exp\left(8\lambda^2 \frac{w_i^2}{\varepsilon_i^2}\right) \qquad \forall |\lambda| < \frac{1}{4\frac{w_i}{\varepsilon_i}}. \tag{26}$$

As such, each $\tilde{Z}_i$, and consequently $\sum_i \tilde{Z}_i$, is a subexponential random variable (Definition 3). In particular, we have that:

$$\sum_{i=1}^n \tilde{Z}_i \in \mathrm{subE}\left(16\sum_{i=1}^n \frac{w_i^2}{\varepsilon_i^2}, 4\max_{i \in [n]} \frac{w_i}{\varepsilon_i}\right). \tag{27}$$

We can therefore apply Fact 1 to show the concentration of the sum of $w_i Z_i$:

$$\mathbf{Pr}\left[\left|\sum_{i=1}^{n} w_i Z_i\right| > \frac{t}{2}\right] \leq 2\exp\left(-\min\left\{\frac{t^2}{128\sum_{i=1}^{n}\frac{w_i^2}{\varepsilon_i^2}}, \frac{t}{16\max_{i\in[n]}\frac{w_i}{\varepsilon_i}}\right\}\right). \tag{28}$$

Combining the tail bounds ((25) and (28)) of both terms in (24), we have:

$$\mathbf{Pr}\left[\left|\hat{\theta}-\theta\right| > t\right] \leq 2\exp\left(-\frac{t^2}{8\sum_{i=1}^{n} w_i^2}\right) + 2\exp\left(-\min\left\{\frac{t^2}{128\sum_{i=1}^{n}\frac{w_i^2}{\varepsilon_i^2}}, \frac{t}{16\max_{i\in[n]}\frac{w_i}{\varepsilon_i}}\right\}\right) \tag{29}$$

$$\leq 4\exp\left(-\frac{1}{8}\min\left\{\frac{t^2}{\sum_{i=1}^{n} w_i^2}, \frac{t^2}{16\sum_{i=1}^{n}\frac{w_i^2}{\varepsilon_i^2}}, \frac{t}{2\max_{i\in[n]}\frac{w_i}{\varepsilon_i}}\right\}\right). \tag{30}$$

We can combine the first and second arguments of the minimum into one argument such that, for some constant c, we have:

$$\mathbf{Pr}\left[\left|\hat{\theta}-\theta\right| > t\right] \leq 4\exp\left(-c\min\left\{\frac{t^2}{\sum_{i=1}^{n} w_i^2\left(1+\frac{1}{\varepsilon_i^2}\right)}, \frac{t}{2\max_{i\in[n]}\frac{w_i}{\varepsilon_i}}\right\}\right). \tag{31}$$

Further, as $\varepsilon_i \leq 1$, we know $w_i/\varepsilon_i \leq w_i(1+1/\varepsilon_i^2)$. This allows us to bring a $(1+1/\varepsilon_i^2)$ term into the second argument, resulting in the expression:

$$\mathbf{Pr}\left[\left|\hat{\theta}-\theta\right| > t\right] \leq 4\exp\left(-c\min\left\{\frac{t^2}{\sum_{i=1}^{n} w_i^2\left(1+\frac{1}{\varepsilon_i^2}\right)}, \frac{t}{2\max_{i\in[n]} w_i\left(1+\frac{1}{\varepsilon_i^2}\right)}\right\}\right). \tag{32}$$

Substituting each w_i for its realization in Algorithm 1, we have:

$$\mathbf{Pr}\left[\left|\hat{\theta}-\theta\right| > t\right] \leq 4\exp\left(-c\sum_{i=1}^{n}\left(1+\frac{1}{\varepsilon_i^2}\right)^{-1}\min\left\{t^2, t\right\}\right). \tag{33}$$

For $\varepsilon_i \leq 1$, $\left(1+\frac{1}{\varepsilon_i^2}\right)^{-1} \approx \varepsilon_i^2$. Incorporating this approximation into (33) leads to:

$$\mathbf{Pr}\left[\left|\hat{\theta}-\theta\right| > t\right] \leq 4\exp\left(-c\sum_{i=1}^{n}\varepsilon_i^2\ \min\left\{t^2, t\right\}\right). \tag{34}$$

Altogether, this implies that with probability at least $1-\beta$,

$$\left|\hat{\theta}-\theta\right|^2 \leq \max\left\{\mathcal{O}\left(\left(\frac{\log(1/\beta)}{\sum_{i=1}^{n}\varepsilon_i^2}\right)^2\right), \mathcal{O}\left(\frac{\log(1/\beta)}{\sum_{i=1}^{n}\varepsilon_i^2}\right)\right\}. \tag{35}$$

When $\mathcal{O}\left(\frac{\log(1/\beta)}{\sum_{i=1}^{n}\varepsilon_i^2}\right) \leq 1$, we achieve the desired bound. If this quantity is greater than 1, outputting $\hat{\theta}=0$ suffices to achieve error 1.

□

B.2 Proof of Theorem 2.2

Theorem 2.2. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0,1)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, $X \in \{-1,1\}$ almost surely. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ and a universal constant c such that, for all $P \in \mathcal{P}$,*

$$\mathbf{Pr}_{X_{1:n}\sim P^{\otimes n}}\left[\left|\hat{\theta}\left(X_{1:n}\right)-\theta\left(P\right)\right|^2 \leq \min\left(c\,\frac{\log(1/\beta)}{\sum_{i=1}^{n}\varepsilon_i^2}, 1\right)\right] \geq 1-\beta. \tag{8}$$

Algorithm 2 achieves this bound.

Proof. Consider Algorithm 1, wherein user i runs ε_i-LDP randomized response on their data, and the server generates a estimate, $\hat{\theta}$, by computing a weighted sum of the users' private data, weighted by $w_i \propto \varepsilon_i^2$ for each user i.

Under this construction, $\mathbb{E}\left[\hat{\theta}\right] = \mathbb{E}\left[\sum_{i=1}^n w_i c_i Y_i\right] = \theta$. Further, as $X_i \in \{-1, 1\}$ for all i, $w_i c_i Y_i \in \{-w_i c_i, w_i c_i\}$. Therefore, we can apply Hoeffding's Inequality to find a high-probability bound for $\left|\hat{\theta} - \theta\right|$:

$$\mathbf{Pr}\left[\left|\hat{\theta} - \theta\right| \geq t\right] \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n w_i^2 c_i^2}\right). \tag{36}$$

Given $w_i \propto 1/c_i^2$ and under the assumption $\varepsilon_i \leq 1$ for all i, we have:

$$\sum_{i=1}^n w_i^2 c_i^2 = \left(\sum_{i=1}^n \frac{1}{c_i^2}\right)^{-1} = \mathcal{O}\left(\left(\sum_{i=1}^n \varepsilon_i^2\right)^{-1}\right). \tag{37}$$

Combining (36) and (37) yields

$$\mathbf{Pr}\left[\left|\hat{\theta} - \theta\right| \geq t\right] \leq 2\exp\left(\mathcal{O}\left(-t^2 \sum_{i=1}^n \varepsilon_i^2\right)\right), \tag{38}$$

implying that with probability at least $1 - \beta$,

$$\left|\hat{\theta} - \theta\right|^2 \leq \mathcal{O}\left(\frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2}\right). \tag{39}$$

Note that if $\mathcal{O}\left(\frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2}\right) \geq 1$, we can simply output $\hat{\theta} = 0$ to achieve error 1.

□

B.3 Proof of Theorem 2.3

Theorem 2.3. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0, 1/2)$. Let $\mathcal{P}$ be the family of distributions P such that for any $X \sim P$, $X \in \{-1, 1\}$ almost surely. For the $\boldsymbol{\varepsilon}$-locally differentially private mean estimation problem over a single dimension, there exists an absolute constant c such that the minimax quantile is lower bounded as*

$$\min\left(c\,\frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2}, 1\right) \leq \mathcal{M}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right). \tag{9}$$

where $c > 0$ is a universal constant.

Proof. Let $\theta \in [-1, 1]$, to be defined later. For all $i \in [n]$, define distributions ${P_1}^{(i)}$ and ${P_2}^{(i)}$ constructively by drawing $X \sim {P_1}^{(i)}$ or $Y \sim {P_2}^{(i)}$ such that:

$$X = \begin{cases} 1 & w.p. \quad \frac{1+\theta}{2} \\ -1 & w.p. \quad \frac{1-\theta}{2} \end{cases} \qquad\qquad Y = \begin{cases} 1 & w.p. \quad \frac{1-\theta}{2} \\ -1 & w.p. \quad \frac{1+\theta}{2} \end{cases}.$$

Let ${Q_1}^{(i)}$ and ${Q_2}^{(i)}$ be the pair of distributions induced by passing ${P_1}^{(i)}$ and ${P_2}^{(i)}$ through an ε_i-differentially private channel, $C^{(i)}$. Formally, define ${Q_1}^{(i)}$ and ${Q_2}^{(i)}$ as

$${Q_1}^{(i)} = C^{(i)}\left({P_1}^{(i)}\right) \quad \text{and} \quad {Q_2}^{(i)} = C^{(i)}\left({P_2}^{(i)}\right). \tag{40}$$

Let Q_1 and Q_2 be the concatenation of the n distributions ${Q_1}^{(1)}, \ldots, {Q_1}^{(n)}$ and ${Q_2}^{(1)}, \ldots, {Q_2}^{(n)}$, respectively. Consider the KL divergence of Q_1 and Q_2. Without loss of generality, assume that $\mathrm{KL}\left(Q_1 \parallel Q_2\right) = \min\left(\mathrm{KL}\left(Q_1 \parallel Q_2\right), \mathrm{KL}\left(Q_2 \parallel Q_1\right)\right)$. As Q_1 and Q_2 are product distributions, their KL divergence is the sum of the divergences of their marginals. This leads to the following bound:

$$\mathrm{KL}\left(Q_1 \parallel Q_2\right) = \min\left(\mathrm{KL}\left(Q_1 \parallel Q_2\right), \mathrm{KL}\left(Q_2 \parallel Q_1\right)\right) \tag{41}$$

$$\leq \frac{1}{2}\left(\mathrm{KL}\left(Q_1 \parallel Q_2\right) + \mathrm{KL}\left(Q_2 \parallel Q_1\right)\right) \tag{42}$$

$$= \frac{1}{2}\sum_{i=1}^{n}\left(\mathrm{KL}\left({Q_1}^{(i)} \,\middle\|\, {Q_2}^{(i)}\right) + \mathrm{KL}\left({Q_2}^{(i)} \,\middle\|\, {Q_1}^{(i)}\right)\right). \tag{43}$$

By Theorem 4, as the distributions induced by ε_i-privacy channels on samples from P_1 and P_2, ${Q_1}^{(i)}$ and ${Q_2}^{(i)}$ satisfy

$$\mathrm{KL}\left({Q_1}^{(i)} \,\middle\|\, {Q_2}^{(i)}\right) + \mathrm{KL}\left({Q_2}^{(i)} \,\middle\|\, {Q_1}^{(i)}\right) \leq \min\left(4, e^{2\varepsilon_i}\right) \cdot \left(e^{\varepsilon_i} - 1\right)^2 \cdot \mathrm{d}_{\mathrm{TV}}^2\left({P_1}^{(i)}, {P_2}^{(i)}\right). \tag{44}$$

By construction, $\mathrm{d}_{\mathrm{TV}}\left({P_1}^{(i)}, {P_2}^{(i)}\right) = \theta$. Combined with the fact that $\varepsilon_i \leq 1$, we can bound (44) as

$$\mathrm{KL}\left({Q_1}^{(i)} \,\middle\|\, {Q_2}^{(i)}\right) + \mathrm{KL}\left({Q_2}^{(i)} \,\middle\|\, {Q_1}^{(i)}\right) \leq 4 \cdot \left(2\varepsilon_i\right)^2 \cdot \theta^2. \tag{45}$$

Combining this bound with (43) gives:

$$\mathrm{KL}\left(Q_1 \parallel Q_2\right) \leq 8\theta^2 \sum_{i=1}^{n} \varepsilon_i^2. \tag{46}$$

To satisfy the assumption of Corollary 1, choose θ as:

$$\theta^2 = \min\left(\frac{\log\left(\frac{1}{4\beta(1-\beta)}\right)}{8\sum_{i=1}^{n}\varepsilon_i^2}, 1\right). \tag{47}$$

Then, $\mathrm{KL}\left(Q_1 \parallel Q_2\right) < \log\left(\frac{1}{4\beta(1-\beta)}\right)$ and we can apply Corollary 1, resulting in a bound for all $\beta \in (0, 1/2)$ of:

$$\mathcal{M}_{-}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right) \geq \frac{1}{2}\left|\theta\left(Q_1\right) - \theta\left(Q_2\right)\right|^2 = \frac{1}{2}\left|\theta + \theta\right|^2 = \theta^2. \tag{48}$$

Finally, by Theorem 5, we can lower bound the minimax quantile $\mathcal{M}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right)$ by the lower minimax quantile $\mathcal{M}_{-}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right)$. Ultimately, we have the following lower bound on $\mathcal{M}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right)$ for $\beta \in (0, 1/2)$:

$$\mathcal{M}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right) \geq \mathcal{M}_{-}\left(\beta, \mathcal{P}, \boldsymbol{\varepsilon}\right) \geq \theta^2 = \min\left(\frac{\log\left(\frac{1}{4\beta(1-\beta)}\right)}{8\sum_{i=1}^{n}\varepsilon_i^2}, 1\right). \tag{49}$$

□

C MISSING PROOFS FROM SECTION 3

C.1 Proof of Theorem 3.1

Theorem 3.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0, 1)$. Let $\mathcal{P}_{2,r}$ be the family of distributions P such that for any $X \sim P$, $X \in \mathbb{B}^d\left(r\right)$ almost surely. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{\theta}$ and a universal constant c such that, for all $P \in \mathcal{P}_{2,r}$,*

$$\mathbf{Pr}_{X_{1:n}\sim P^{\otimes n}}\left[\left\|\hat{\theta}\left(X_{1:n}\right) - \theta\left(P\right)\right\|_2^2 \leq r^2 \min\left(c\frac{\left(d + \log\left(1/\beta\right)\right)}{\sum_{i=1}^{n}\varepsilon_i^2}, 1\right)\right] \geq 1 - \beta. \tag{10}$$

Algorithm 3 achieves this bound.

Proof. Consider Algorithm 3, wherein each user i generates an ε_i-LDP estimate Y_i of their data point X_i according to Algorithm 4, and the server outputs a weighted average $\hat{\theta} = \sum_i w_i Y_i$. Note that Algorithm 4

produces an unbiased estimate of X_i and recall that $\mathbb{E}[X_i] = \theta$. Applying Cauchy-Schwarz and a union bound, we can bound the tail of this algorithm as:

$$\mathbf{Pr}\left[\left\|\hat{\theta} - \theta\right\|_2 \leq t\right] \geq \mathbf{Pr}\left[\left\|\sum_{i=1}^n w_i Y_i - \sum_{i=1}^n w_i X_i\right\|_2 \leq \frac{t}{2}, \left\|\sum_{i=1}^n w_i X_i - \theta\right\|_2 \leq \frac{t}{2}\right]. \tag{50}$$

It follows from the law of total probability that:

$$\mathbf{Pr}\left[\left\|\hat{\theta} - \theta\right\|_2 \leq t\right] \geq \mathbf{Pr}\left[\left\|\sum_{i=1}^n w_i Y_i - \sum_{i=1}^n w_i X_i\right\|_2 \leq \frac{t}{2} \mid X_1, \ldots, X_n\right] \cdot \mathbf{Pr}\left[\left\|\sum_{i=1}^n w_i X_i - \theta\right\|_2 \leq \frac{t}{2}\right]. \tag{51}$$

We will establish tail bounds on each of these terms separately.

First term of (51). First, consider the first term. We begin by showing that, for all i, $Y_i - X_i$ conditioned on X_i is a subgaussian random vector with variance proxy independent of d. Equivalently, we show that for any $\ell \in \mathbb{S}^{d-1}$, $\langle \ell, Y_i - X_i \rangle$ is subgaussian.

Fix $i \in [n]$. Without loss of generality, let $X_i = \mathbf{e}_d$. Let $S_1 = \{Y \in \mathbb{S}^{d-1}(B_i) \text{ s.t. } \langle Y, X_i \rangle > 0\}$, and let $S_2 = S_1^{\complement} = \{Y \in \mathbb{S}^{d-1}(B_i) \text{ s.t. } \langle Y, X_i \rangle \leq 0\}$. Let Q_1 be the uniform distribution over S_1, and let Q_2 be the uniform distribution over S_2. Denote by Q the distribution of Y_i. Then, Q is a mixture of Q_1 and Q_2.

Let p be the probability that $Y_i \sim Q_1$, and let $1-p$ be the probability that $Y_i \sim Q_2$. The event $Y_i \sim Q_1$ can occur in two ways: first, in Line 1 of Algorithm 3, we choose $\tilde{X}_i \propto -X_i$ and then, in Line 8, we have $T = 0$; or, in Line 1 of Algorithm 3, we choose $\tilde{X}_i \propto X_i$ and then, in Line 8, we have $T = 1$. Given the probabilities of each of these events, we can find p to be:

$$p = \frac{1}{2} + \frac{\|X_i\|_2}{2r}\left(\frac{e_i^\varepsilon - 1}{e_i^\varepsilon + 1}\right). \tag{52}$$

Define a mirroring function $\textsc{mirror} : S_2 \to S_1$ as:

$$\textsc{mirror}\left((x_1, \ldots, x_{d-1}, x_d)\right) = (x_1, \ldots, x_{d-1}, -x_d). \tag{53}$$

Now, for an arbitrary $\ell \in \mathbb{S}^{d-1}$, let $f : \mathbb{S}^{d-1}(B_i) \to \mathbb{R}$ be such that $f(Y) = \langle \ell, Y \rangle$. By construction, f is a 1-Lipschitz function over the sphere of radius B_i. Similarly, define $g : \mathbb{S}^{d-1}(B_i) \to \mathbb{R}$ as follows:

$$g(Y) = \begin{cases} f(Y) & Y \in S_1 \\ f(\textsc{mirror}(Y)) & Y \in S_2 \end{cases}. \tag{54}$$

Because f is 1-Lipschitz, g is also 1-Lipschitz. If two points lie on the same hemisphere, their distance under g is the same as under f, whereas if they lie on different hemispheres, the mirroring operation brings them closer together, so their distance under g is less than that under f.

We will now prove that the distribution of $f - \mathbb{E}_Q[f]$ has subgaussian tail bounds. If Q was uniform over the entire sphere, we could apply Levy's Lemma immediately to establish this result. However, Q is a mixture of two uniform distributions over the sphere. To prove subgaussianity, we will show that the distribution of $f - \mathbb{E}_{Q_1}[f]$ over Q_1 is subgaussian, re-express Q as a mixture of the uniform distribution and Q_1, and prove that, because of the subgaussianity of its parts, $f - \mathbb{E}_Q[f]$ over Q is subgaussian.

Subgaussianity over Q_1. Let **Unif** denote the uniform distribution over $\mathbb{S}^{d-1}(B_i)$. We begin by relating the distribution of $f - \mathbb{E}_{Q_1}[f]$ over S_1 with the distribution of $g - \mathbb{E}_{\mathbf{Unif}}[g]$ over **Unif**. First, because **Unif** can be expressed as an equal mixture of Q_1 and Q_2, we have:

$$\mathbf{Pr}_{Y \sim \mathbf{Unif}}[|g(Y) - \mathbb{E}_{Y \sim \mathbf{Unif}}[g(Y)]|] \tag{55}$$

$$= \frac{1}{2}\mathbf{Pr}_{Y \sim Q_1}[|g(Y) - \mathbb{E}_{Y \sim \mathbf{Unif}}[g(Y)]|] + \frac{1}{2}\mathbf{Pr}_{Y \sim Q_2}[|g(Y) - \mathbb{E}_{Y \sim \mathbf{Unif}}[g(Y)]|]. \tag{56}$$

For $Y \sim Q_1$, we know $g(Y) = f(Y)$. Additionally, we can find the expectation of g over **Unif** to be:

$$\mathbb{E}_{Y \sim \mathbf{Unif}}[g(Y)] = \frac{1}{2}\mathbb{E}_{Y \sim Q_1}[g(Y)] + \frac{1}{2}\mathbb{E}_{Y \sim Q_2}[g(Y)] \tag{57}$$

$$= \frac{1}{2}\mathbb{E}_{Y\sim Q_1}[f(Y)] + \frac{1}{2}\mathbb{E}_{Y\sim Q_2}[f(\textsc{mirror}(Y))] \tag{58}$$

$$= \frac{1}{2}\mathbb{E}_{Y\sim Q_1}[f(Y)] + \frac{1}{2}\mathbb{E}_{Y\sim Q_1}[f(Y)] \tag{59}$$

$$= \mathbb{E}_{Y\sim Q_1}[f(Y)]. \tag{60}$$

Combining these facts with (56), we have:

$$\mathbf{Pr}_{Y\sim\mathbf{Unif}}[|g(Y) - \mathbb{E}_{Y\sim\mathbf{Unif}}[g(Y)]|] \tag{61}$$

$$= \frac{1}{2}\mathbf{Pr}_{Y\sim Q_1}[|f(Y) - \mathbb{E}_{Y\sim Q_1}[f(Y)]|] + \frac{1}{2}\mathbf{Pr}_{Y\sim Q_2}[|g(Y) - \mathbb{E}_{Y\sim\mathbf{Unif}}[g(Y)]|] \tag{62}$$

$$= \frac{1}{2}\mathbf{Pr}_{Y\sim Q_1}[|f(Y) - \mathbb{E}_{Y\sim Q_1}[f(Y)]|] + \frac{1}{2}\mathbf{Pr}_{Y\sim Q_1}[|f(Y) - \mathbb{E}_{Y\sim\mathbf{Unif}}[f(Y)]|] \tag{63}$$

$$= \mathbf{Pr}_{Y\sim Q_1}[|f(Y) - \mathbb{E}_{Y\sim Q_1}[f(Y)]|]. \tag{64}$$

Because g is 1-Lipschitz over $\mathbb{S}^{d-1}(B_i)$, we can apply Levy's Lemma (Theorem 3.1), resulting in:

$$\mathbf{Pr}_{Y\sim Q_1}[|f(Y) - \mathbb{E}_{Y\sim Q_1}[f(Y)]|] \le 4\exp\left(-\frac{2Cdt^2}{B_i^2}\right). \tag{65}$$

Therefore, $f(Y) - \mathbb{E}_{Y\sim Q_1}[f(Y)]$ is $\mathcal{O}\left(\frac{B_i^2}{d}\right)$-subgaussian over Q_1. Similarly, by Levy's Lemma (Theorem 3.1),

$$\mathbf{Pr}_{Y\sim\mathbf{Unif}}[|f(Y) - \mathbb{E}_{Y\sim\mathbf{Unif}}[f(Y)]|] \le 2\exp\left(-\frac{2Cdt^2}{B_i^2}\right), \tag{66}$$

implying that $f(Y) - \mathbb{E}_{Y\sim\mathbf{Unif}}[f(Y)]$ is also $\mathcal{O}\left(\frac{B_i^2}{d}\right)$-subgaussian over **Unif**.

Subgaussianity over Q. Recall that Q is a mixture of Q_1 and Q_2:

$$Q = p\cdot Q_1 + (1-p)\cdot Q_2. \tag{67}$$

However, because **Unif** is also an equal mixture of Q_1 and Q_2, we could alternatively express Q as a mixture of Q_1 and **Unif** as follows:

$$Q = (2p-1)\cdot Q_1 + 2(1-p)\cdot \mathbf{Unif}. \tag{68}$$

We continue to prove that $f - \mathbb{E}_Q[f]$ is subgaussian by proving that it satisfies the second property described in Definition 4.

Manipulating the expression given by the second property of Definition 4 for **Unif** and Q_1, respectively, we have, for some constants c_1 and c_2:

$$\mathbb{E}_{Y\sim\mathbf{Unif}}[\exp(t\cdot f(Y))] \le \exp\left(\mathcal{O}\left(\frac{B_i^2t^2}{d}\right)\right)\exp(t\cdot\mathbb{E}_{Y\sim\mathbf{Unif}}[f(Y)]) \tag{69}$$

$$\mathbb{E}_{Y\sim Q_1}[\exp(t\cdot f(Y))] \le \exp\left(\mathcal{O}\left(\frac{B_i^2t^2}{d}\right)\right)\exp(t\cdot\mathbb{E}_{Y\sim Q_1}[f(Y)]). \tag{70}$$

Applying these bounds and the law of total expectation yields, for any $t\in\mathbb{R}$,

$$\mathbb{E}_{Y\sim Q}[\exp(t(f(Y) - \mathbb{E}_{Y\sim Q}[f(Y)]))] \tag{71}$$

$$\le \exp(-t\,\mathbb{E}_{Y\sim Q}[f(Y)])\left(2(1-p)\exp\left(\mathcal{O}\left(\frac{B_i^2t^2}{d}\right)\right)\exp(t\cdot\mathbb{E}_{Y\sim\mathbf{Unif}}[f(Y)])\right. \tag{72}$$

$$\left. +(2p-1)\exp\left(\mathcal{O}\left(\frac{B_i^2t^2}{d}\right)\right)\exp(t\cdot\mathbb{E}_{Y\sim Q_1}[f(Y)])\right). \tag{73}$$

As f represents an inner product between B_iY and ℓ and the expected value of Y drawn uniformly from the sphere is the zero vector, we have $\mathbb{E}_{Y\sim\mathbf{Unif}}[f(Y)] = 0$. Additionally, conditioned on X_i, the expected value of Y drawn from Q is X_i (Duchi et al., 2013). Therefore,

$$f(X_i) = \mathbb{E}_{Y\sim Q}[f(Y)] = (2p-1)\mathbb{E}_{Y\sim Q_1}[f(Y)], \tag{74}$$

implying $\mathbb{E}_{Y\sim Q_1}[f(Y)] = \langle \ell, X_i/(2p-1)\rangle$. Recall that p was the probability that Y_i was drawn from the hemisphere on which $\langle Y_i, X_i\rangle > 0$, described in (52). Substituting for p gives us a bound on (73) of:

$$\mathbb{E}_{Y\sim Q}[\exp(t(f(Y) - \mathbb{E}_{Y\sim Q}[f(Y)]))] \tag{75}$$

$$\leq \exp(-t\,\langle \ell, X_i\rangle)\exp\left(\mathcal{O}\left(\frac{B_i^2 t^2}{d}\right)\right)\cdot\left(1+\exp\left(tr\left(\frac{e_i^\varepsilon+1}{e_i^\varepsilon-1}\right)\frac{\langle \ell, X_i\rangle}{\|X_i\|_2}\right)\right). \tag{76}$$

Taking absolute values, along with the facts that $\frac{e_i^\varepsilon+1}{e_i^\varepsilon-1} = \mathcal{O}(1/\varepsilon_i)$ and $|\langle \ell, X_i\rangle / \|X_i\|_2| \leq 1$ leaves us with:

$$\begin{aligned}\mathbb{E}_{Y\sim Q}[\exp(t(f(Y) - \mathbb{E}_{Y\sim Q}[f(Y)]))] &\leq \exp\left(\mathcal{O}\left(\frac{B_i^2t^2}{d}\right)\right)\cdot\left(\exp(|t|\,r)+\exp\left(\frac{|t|\,r}{\varepsilon_i}\right)\right)\\ &\leq 2\exp\left(\mathcal{O}\left(\frac{B_i^2t^2}{d}\right)\right)\cdot\left(\exp\left(\frac{|t|\,r}{\varepsilon_i}\right)\right).\end{aligned}$$

Given $B_i^2 \leq r^2 d/\varepsilon_i$, this is:

$$\mathbb{E}_{Y\sim Q}[\exp(t(f(Y) - \mathbb{E}_{Y\sim Q}[f(Y)]))] \leq 2\exp\left(\mathcal{O}\left(\frac{r^2t^2}{\varepsilon_i^2}\right)\right)\cdot\left(\exp\left(\frac{|t|\,r}{\varepsilon_i}\right)\right) \tag{77}$$

$$\leq 2e\cdot\exp\left(2\mathcal{O}\left(\frac{r^2t^2}{\varepsilon_i^2}\right)\right). \tag{78}$$

Therefore, by Definition 4, $f(Y_i) - \mathbb{E}_{Y_i\sim Q}[f(Y_i)] = \langle \ell, Y_i - X_i\rangle$ is $\mathcal{O}\left(r^2/\varepsilon_i^2\right)$-subgaussian for all i, ℓ.

Error bound from subgaussianity.

The properties of subgaussianity and the fact that $\sum_i w_i = 1$ then imply that $\langle \ell, \sum_{i=1}^n w_i(Y_i - X_i)\rangle$ is itself subgaussian, with variance proxy at most $\mathcal{O}\left(r^2\sum_{i=1}^n \frac{w_i^2}{\varepsilon_i^2}\right)$. Therefore, $\sum_{i=1}^n w_i(Y_i - X_i)$ is an $\mathcal{O}\left(r^2\sum_{i=1}^n \frac{w_i^2}{\varepsilon_i^2}\right)$-subgaussian random vector. As such, we can apply Fact 2 to find tail bounds on its norm, therefore implying that, conditioned on $X_i, \ldots, X_n$, with probability at least $1-\beta/2$,

$$\left\|\sum_{i=1}^n w_i(Y_i - X_i)\right\|_2 \leq \mathcal{O}\left(\sqrt{r^2\sum_{i=1}^n \frac{w_i^2}{\varepsilon_i^2}\left(d+\log\left(\frac{2}{\beta}\right)\right)}\right). \tag{79}$$

Finally, as each w_i is chosen proportionally to ε_i^2, with probability at least $1-\beta/2$, conditioned on $X_i, \ldots, X_n$,

$$\left\|\sum_{i=1}^n w_i(Y_i - X_i)\right\|_2 \leq \mathcal{O}\left(\sqrt{\frac{r^2(d+\log(2/\beta))}{\sum_{i=1}^n \varepsilon_i^2}}\right). \tag{80}$$

Second term of (50). To bound the second term of (50), we apply a Hoeffding-like bound for norm-subgaussian random variables (Jin et al., 2019). First, note that, as $\sum_i w_i = 1$, we can express this term as the norm of a sum of centered random vectors $Z_i := w_i X_i - \mathbb{E}[w_i X_i]$:

$$\left\|\sum_{i=1}^n w_i X_i - \theta\right\|_2 = \left\|\sum_{i=1}^n w_i X_i - \mathbb{E}[w_i X_i]\right\|_2 = \left\|\sum_{i=1}^n Z_i\right\|_2. \tag{81}$$

Each of these random variables has a bounded norm. Specifically, $\|Z_i\|_2 \leq 2w_i r$. This implies that Z_i is $4w_i^2r^2$-norm-subgaussian by Fact 3 . Therefore, by Fact 4, with probability at least $1-\beta/2$,

$$\left\|\sum_{i=1}^n w_i X_i - \theta\right\|_2 = \mathcal{O}\left(\sqrt{r^2\log(2d/\beta)\sum_{i=1}^n w_i^2}\right). \tag{82}$$

Plugging in $w_i = \varepsilon_i^2/\sum_j \varepsilon_j^2$, we have, with probability at least $1-\beta/2$,

$$\left\|\sum_{i=1}^n w_i X_i - \theta\right\|_2 \leq \mathcal{O}\left(\sqrt{\frac{r^2\log(2d/\beta)}{\sum_{i=1}^n \varepsilon_i^2}}\right) \leq \mathcal{O}\left(\sqrt{\frac{r^2(d+\log(2/\beta))}{\sum_{i=1}^n \varepsilon_i^2}}\right). \tag{83}$$

Both terms. Setting $\frac{t}{2}$ in (51) to be $\mathcal{O}\left(\sqrt{\frac{r^2(d+\log(2/\beta))}{\sum_{i=1}^n \varepsilon_i^2}}\right)$, we have by (80) and (83),

$$\mathbf{Pr}\left[\left\|\sum_{i=1}^n w_i Y_i - \sum_{i=1}^n w_i X_i\right\|_2 \le \frac{t}{2} \mid X_1, \ldots, X_n\right] \cdot \mathbf{Pr}\left[\left\|\sum_{i=1}^n w_i X_i - \theta\right\|_2 \le \frac{t}{2}\right] \ge \left(1 - \frac{\beta}{2}\right)^2 \ge 1 - \beta. \tag{84}$$

Combined with (51), this implies that with probability at least $1-\beta$,

$$\left\|\hat{\theta} - \theta\right\|_2 \le \mathcal{O}\left(\sqrt{\frac{r^2\left(d + \log\left(1/\beta\right)\right)}{\sum_{i=1}^n \varepsilon_i^2}}\right). \tag{85}$$

An upper bound of r^2 can be achieved by outputting the zero vector.

□

C.2 Proof of Theorem 3.2

Theorem 3.2. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \le 1$ for all i. Let $\beta \in (0, 3/20)$. Let $r > 0$. For all $i \in [n]$, let $\varepsilon_i \in (0,1)$. For the mean estimation problem over the ℓ_2 ball with radius r, there exists an absolute constant c such that the minimax quantile is lower bounded as*

$$\mathcal{M}\left(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}\right) \ge c\, r^2 \min\left(\frac{\log(1/\beta) + d}{\sum_{i=1}^n \varepsilon_i^2}, \frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2} + \frac{1}{\sqrt{\sum_{i=1}^n \varepsilon_i^2}}, 1\right).$$

Proof. By Theorem 2.3, we have

$$\mathcal{M}\left(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}\right) \ge r^2 \min\left(\log\left(\frac{1}{\beta}\right), 1\right). \tag{86}$$

To establish the remaining terms, we begin by proving a lower bound on the minimax expected error of the multi-dimensional mean estimation problem. Specifically, for all $k \in [d]$, we will construct $\mathcal{P}^{(k)} \subseteq \mathcal{P}_{2,r}$ such that:

$$\inf_{Q \in \mathcal{Q}_{\boldsymbol{\varepsilon}}} \inf_{\hat{\theta} \in \hat{\Theta}} \sup_{P_\nu \in \mathcal{P}^{(k)}} \mathbb{E}_{P_\nu, Q}\left[\left\|\hat{\theta}\left(Z_{1:n}\right) - \theta\left(P_\nu\right)\right\|_2^2\right] \ge \Theta\left(r^2 \min\left(\frac{1}{k}, \frac{k}{\sum_{i=1}^n \varepsilon_i^2}\right)\right). \tag{87}$$

To do so, let $\psi \in (0,1]$, to be specified later in the proof. Let $k \in [d]$. Construct $\mathcal{P}^{(k)} \subseteq \mathcal{P}_{2,r}$ as follows. Let $\mathcal{V}_k = \{-1,1\}^k$. For all $\nu \in \mathcal{V}_k$ and for all $i \in [n]$, define a distribution $P_\nu^{(i)} \in \mathcal{P}^{(k)}$ constructively by first drawing $j \sim \mathbf{Unif}\left([k]\right)$, then drawing X such that

$$X = \begin{cases} r\mathbf{e}_j & \text{w.p.} \frac{1+\psi\nu_j}{2} \\ -r\mathbf{e}_j & \text{w.p.} \frac{1-\psi\nu_j}{2} \end{cases},$$

where $\mathbf{e}_j$ is the j^{th} basis vector in $\mathbb{R}^d$. For all $\nu \in \mathcal{V}_k$, let $Q_\nu^{(i)}$ be the distribution induced by passing $P_\nu^{(i)}$ through an ε_i-differentially private channel, $C^{(i)}$. Formally, $Q_\nu^{(i)} = C^{(i)}\left(P_\nu^{(i)}\right)$. Let Q_ν be the concatenation of the n distributions ${Q_\nu}^{(1)}, \ldots, {Q_\nu}^{(n)}$. For all $\nu, \nu' \in \mathcal{V}_k$, we write $\nu \sim \nu'$ if ν and ν' differ in exactly one coordinate. If ν and ν' differ only in coordinate j, we write $\nu \sim_j \nu'$.

Then, under Assouad's Lemma (Lemma 1), we have

$$\begin{aligned} \inf_{Q \in \mathcal{Q}_{\boldsymbol{\varepsilon}}} \inf_{\hat{\theta} \in \hat{\Theta}} \sup_{P_\nu \in \mathcal{P}^{(k)}} \mathbb{E}_{P_\nu, Q}&\left[\sum_{j=1}^k \left(\hat{\theta}\left(Z_{1:n}\right)_j - \theta\left(P_\nu\right)_j\right)^2\right] \\ &\ge \frac{k}{2} \min_{j \in [k], \nu \sim_j \nu'} \left(\theta\left(Q_\nu\right)_j - \theta\left(Q_{\nu'}\right)_j\right)^2 \min_{\nu \sim \nu'} \left(1 - \mathrm{d_{TV}}\left(Q_\nu, Q_{\nu'}\right)\right). \end{aligned} \tag{88}$$

Consider the total variation distance between Q_ν and $Q_{\nu'}$ for $\nu \sim \nu'$. By Pinsker's Inequality, we can bound the square of this total variation distance by the KL-divergence of Q_ν and $Q_{\nu'}$ as follows:

$$\begin{aligned} \mathrm{d}_{\mathrm{TV}}^2\left(Q_\nu, Q_{\nu'}\right) &\leq \frac{1}{2}\min\left(\mathrm{KL}\left(Q_\nu \parallel Q_{\nu'}\right), \mathrm{KL}\left(Q_{\nu'} \parallel Q_\nu\right)\right) && (89)\\ &\leq \frac{1}{4}\left(\mathrm{KL}\left(Q_\nu \parallel Q_{\nu'}\right) + \mathrm{KL}\left(Q_{\nu'} \parallel Q_\nu\right)\right) && (90)\\ &= \frac{1}{4}\sum_{i=1}^{n}\left(\mathrm{KL}\left(Q_\nu^{(i)} \,\middle\|\, Q_{\nu'}^{(i)}\right) + \mathrm{KL}\left(Q_{\nu'}^{(i)} \,\middle\|\, Q_\nu^{(i)}\right)\right), && (91)\end{aligned}$$

where the equality is due to Q_ν and $Q_{\nu'}$ being product distributions. By Theorem 4, as the distributions induced by ε_i-privacy channels on samples from $P_\nu^{(i)}$ and $P_{\nu'}^{(i)}$, $Q_\nu^{(i)}$ and $Q_{\nu'}^{(i)}$ satisfy:

$$\mathrm{KL}\left(Q_\nu^{(i)} \,\middle\|\, Q_{\nu'}^{(i)}\right) + \mathrm{KL}\left(Q_{\nu'}^{(i)} \,\middle\|\, Q_\nu^{(i)}\right) \leq \min\left(4, e^{2\varepsilon_i}\right)\cdot\left(e^{\varepsilon_i}-1\right)^2\cdot \mathrm{d}_{\mathrm{TV}}^2\left(P_\nu^{(i)}, P_{\nu'}^{(i)}\right). \tag{92}$$

By the construction of $P_\nu^{(i)}$ and $P_{\nu'}^{(i)}$ and given that ν and ν' differ in exactly one coordinate, $\mathrm{d}_{\mathrm{TV}}\left(P_\nu^{(i)}, P_{\nu'}^{(i)}\right) = \psi/k$. Combined with the fact that $\varepsilon_i \leq 1$, we can bound (92) as

$$\mathrm{KL}\left(Q_\nu^{(i)} \,\middle\|\, Q_{\nu'}^{(i)}\right) + \mathrm{KL}\left(Q_{\nu'}^{(i)} \,\middle\|\, Q_\nu^{(i)}\right) \leq \frac{4\cdot\left(2\varepsilon_i\right)^2\cdot\psi^2}{k^2}. \tag{93}$$

Combining this bound with (91) gives:

$$\mathrm{d}_{\mathrm{TV}}^2\left(Q_\nu, Q_{\nu'}\right) \leq 4\frac{\psi^2}{k^2}\sum_{i=1}^{n}\varepsilon_i^2. \tag{94}$$

Choose $\psi^2 = \min\left(k^2/\left(16\sum_{i=1}^n \varepsilon_i^2\right), 1\right)$. Then, $\mathrm{d}_{\mathrm{TV}}\left(Q_\nu, Q_{\nu'}\right) \leq 1/2$. As a result, we can lower bound (88) as:

$$\inf_{Q\in\mathcal{Q}_{\boldsymbol{\varepsilon}}}\inf_{\hat{\theta}\in\hat{\Theta}}\sup_{P_\nu\in\mathcal{P}^{(k)}}\mathbb{E}_{P_\nu,Q}\left[\sum_{j=1}^{k}\left(\hat{\theta}\left(Z_{1:n}\right)_j - \theta\left(P_\nu\right)_j\right)^2\right] \geq \frac{k}{4}\min_{j\in[k],\nu\sim_j\nu'}\left(\theta\left(Q_\nu\right)_j - \theta\left(Q_{\nu'}\right)_j\right)^2, \tag{95}$$

Consider now, for all $j\in[k]$, $\min_{\nu\sim_j\nu'}\left(\theta\left(Q_\nu\right)_j - \theta\left(Q_{\nu'}\right)_j\right)^2$. For all ν, by the construction of Q_ν, $\theta\left(Q_\nu\right)_j = \mathbb{E}[Q_\nu]_j = \frac{\psi r}{k}\nu_j$. Given Q_ν and $Q_{\nu'}$ such that $\nu\sim_j\nu'$, we have:

$$\left(\theta\left(Q_\nu\right)_j - \theta\left(Q_{\nu'}\right)_j\right)^2 = \left\langle \mathbf{e}_j, \theta_{\nu,j} - \theta_{\nu',j}\right\rangle^2 = \frac{\psi^2 r^2}{k^2}\left\langle \mathbf{e}_j, \nu-\nu'\right\rangle^2 = \frac{4\psi^2 r^2}{k^2}. \tag{96}$$

Substituting this equality into (95) leads to:

$$\inf_{Q\in\mathcal{Q}_{\boldsymbol{\varepsilon}}}\inf_{\hat{\theta}\in\hat{\Theta}}\sup_{P_\nu\in\mathcal{P}^{(k)}}\mathbb{E}_{P_\nu,Q}\left[\sum_{j=1}^{k}\left(\hat{\theta}\left(Z_{1:n}\right)_j - \theta\left(P_\nu\right)_j\right)^2\right] \geq \frac{\psi^2 r^2}{k}. \tag{97}$$

Ultimately, as the sum inside the expectation of (97) is upper bounded by the ℓ_2^2-norm of $\hat{\theta}-\theta(P_\nu)$, we have a lower bound on the minimax risk for the class $\mathcal{P}^{(k)}$, given by:

$$\inf_{Q\in\mathcal{Q}_{\boldsymbol{\varepsilon}}}\inf_{\hat{\theta}\in\hat{\Theta}}\sup_{P_\nu\in\mathcal{P}^{(k)}}\mathbb{E}_{P_\nu,Q}\left[\left\|\hat{\theta}\left(Z_{1:n}\right) - \theta\left(P_\nu\right)\right\|_2^2\right] \geq \inf_{Q\in\mathcal{Q}_{\boldsymbol{\varepsilon}}}\inf_{\hat{\theta}\in\hat{\Theta}}\sup_{P_\nu\in\mathcal{P}^{(k)}}\mathbb{E}_{P_\nu,Q}\left[\sum_{j=1}^{k}\left(\hat{\theta}\left(Z_{1:n}\right)_j - \theta\left(P_\nu\right)_j\right)^2\right] \geq \frac{\psi^2 r^2}{k}. \tag{98}$$

To translate this local minimax risk bound into a bound on the related minimax quantile, we will apply Theorem 6. Let $D^2 = \sup_{P_\nu, P_{\nu'}\in\mathcal{P}^{(k)}}\left\|\theta\left(P_\nu\right) - \theta\left(P_{\nu'}\right)\right\|^2$. We can calculate D^2 as:

$$D^2 = \sup_{P_\nu, P_{\nu'}\in\mathcal{P}^{(k)}}\left\|\theta\left(P_\nu\right) - \theta\left(P_{\nu'}\right)\right\|_2^2 = \frac{\psi^2 r^2}{k^2}\sup_{\nu,\nu'\in\mathcal{V}_k}\left\|\nu-\nu'\right\|_2^2 = \frac{4\psi^2 r^2}{k} \tag{99}$$

Therefore, by Theorem 6 with $\varphi = 1/4$, we have, for all $\beta \in \left(0, \frac{3}{20}\right)$,

$$\mathcal{M}_{-}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \geq \frac{\psi^2 r^2}{4k}. \tag{100}$$

By Theorem 5, we can lower bound the minimax quantile $\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon})$ by the lower minimax quantile $\mathcal{M}_{-}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon})$, yielding:

$$\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \geq \mathcal{M}_{-}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \geq \frac{\psi^2 r^2}{4k}. \tag{101}$$

Substituting for $\psi^2 = \min\left(1, \frac{k^2}{16\sum_{i=1}^n \varepsilon_i^2}\right)$, we have

$$\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \geq \frac{r^2}{4} \min\left(\frac{1}{k}, \frac{k}{16\sum_{i=1}^n \varepsilon_i^2}\right). \tag{102}$$

Given that (102) holds for all $k \in [d]$, we can lower bound $\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon})$ by a maximization over all k as follows:

$$\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \geq \frac{r^2}{4} \max_{k \in [d]} \min\left(\frac{1}{k}, \frac{k}{16\sum_{i=1}^n \varepsilon_i^2}\right) = \frac{r^2}{4} \min\left(1, \frac{d}{16\sum_{i=1}^n \varepsilon_i^2}, \frac{1}{4\sqrt{\sum_{i=1}^n \varepsilon_i^2}}\right). \tag{103}$$

Finally, combining Theorem 2.3 and (103), we find, for some constant $c_2 > 0$,

$$\mathcal{M}(\beta, \mathcal{P}_{2,r}, \boldsymbol{\varepsilon}) \geq c_2 \cdot r^2 \min\left(1, \frac{\log(1/\beta) + d}{\sum_{i=1}^n \varepsilon_i^2}, \frac{\log(1/\beta)}{\sum_{i=1}^n \varepsilon_i^2} + \frac{1}{\sqrt{\sum_{i=1}^n \varepsilon_i^2}}\right). \tag{104}$$

□

D MISSING PROOFS FROM SECTION 4

D.1 Proof of Theorem 4.1

Theorem 4.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0,1)$. Let $\mathcal{P}$ be the family of distributions P over $[d]$. There exists an $\boldsymbol{\varepsilon}$-locally differentially private estimator $\hat{P}$ and a universal constant c such that, for all $P \in \mathcal{P}$,*

$$\mathbf{Pr}_{x_{1:n} \sim P^{\otimes n}}\left[\left\|\hat{P}(x_{1:n}) - P\right\|_\infty \leq \min\left(c\sqrt{\frac{\log(d/\beta)}{\sum_{i=1}^n \varepsilon_i^2}}, 1\right)\right] \geq 1 - \beta. \tag{11}$$

Algorithm 5 achieves this bound.

The proof of Theorem 4.1 relies on the following claim.

Lemma D.1. *Let $\boldsymbol{\varepsilon} = \{\varepsilon_i\}_{i=1}^n$, with $\varepsilon_i \leq 1$ for all i. Let $\beta \in (0,1)$. Fix $\boldsymbol{\Phi} \in \{-1\sqrt{m}, 1\sqrt{m}\}^{m \times d}$. Let $v \in [d]$. Let $\bar{\mathbf{z}}$ be the result of Line 10 of Algorithm 5. Then, for $v \in [d]$, $\bar{\mathbf{z}}$ satisfies:*

$$\mathbf{Pr}\left[|\langle \bar{\mathbf{z}} - \mathbb{E}[\bar{\mathbf{z}}], \boldsymbol{\Phi}\mathbf{e}_v\rangle| \leq \sqrt{2\log\left(\frac{2}{\beta}\right) \cdot \left(\sum_{i=1}^n \frac{1}{c_i^2}\right)^{-1}}\right] \geq 1 - \beta, \tag{105}$$

where the probability is taken over the randomness in sampling j and $\mathbf{z}_i$ for each user.

Proof of Claim D.1. Rewriting $\bar{\mathbf{z}}$ as $\sum_i w_i \mathbf{z}_i$, we have:

$$\begin{aligned}
|\langle \bar{\mathbf{z}} - \mathbb{E}[\bar{\mathbf{z}}], \boldsymbol{\Phi}\mathbf{e}_v\rangle| &= \left|\left\langle \sum_{i=1}^n w_i \mathbf{z}_i - \mathbb{E}\left[\sum_{i=1}^n w_i \mathbf{z}_i\right], \boldsymbol{\Phi}\mathbf{e}_v \right\rangle\right| \\
&= \left|\left\langle \sum_{i=1}^n w_i \mathbf{z}_i, \boldsymbol{\Phi}\mathbf{e}_v \right\rangle - \left\langle \mathbb{E}\left[\sum_{i=1}^n w_i \mathbf{z}_i\right], \boldsymbol{\Phi}\mathbf{e}_v \right\rangle\right|
\end{aligned}$$

$$= \left| \sum_{i=1}^{n} w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}\left[\sum_{i=1}^{n} w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle \right] \right|.$$

We have that, for each i,

$$\langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle = \sum_{j=1}^{d} \mathbf{z}_{ij} \mathbf{\Phi}_{jv}.$$

Each $\mathbf{z}_i$ is non-zero for one index $j' \in [d]$. This implies that $\langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle = \mathbf{z}_{ij'} \mathbf{\Phi}_{j'v}$. By construction, $\mathbf{z}_i \in \{-c_i\sqrt{d}, 0, c_i\sqrt{d}\}^m$ for all i and $\mathbf{\Phi} \in \{-1/\sqrt{m}, 1/\sqrt{m}\}^{m \times d}$. As a result, we know that $\langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle \in \{-c_i, c_i\}$ and $w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle \in \{-w_i c_i, w_i c_i\}$. We can thus apply a standard Hoeffding bound and observe that, for all $t > 0$,

$$\mathbf{Pr}\left[\left| \sum_{i=1}^{n} w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}\left[\sum_{i=1}^{n} w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle \right] \right| \geq t \right] \leq 2 \exp\left(- \frac{t^2}{2 \sum_{i=1}^{n} w_i^2 c_i^2} \right). \tag{106}$$

As each w_i is chosen proportionally to $1/c_i^2$, (106) is:

$$\mathbf{Pr}\left[\left| \sum_{i=1}^{n} w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}\left[\sum_{i=1}^{n} w_i \langle \mathbf{z}_i, \mathbf{\Phi} \mathbf{e}_v \rangle \right] \right| \geq t \right] \leq 2 \exp\left(- \frac{t^2}{2} \sum_{i=1}^{n} \frac{1}{c_i^2} \right). \tag{107}$$

Setting $t^2 = 2 \log\left(\frac{2}{\beta}\right) \cdot \left(\sum_{i=1}^{n} \frac{1}{c_i^2} \right)^{-1}$ in (106) completes the proof. □

Given Claim D.1, we can now prove Theorem 4.1.

Proof of Theorem 4.1. For all values $v \in [d]$, let $P(v)$ be the true probability of observing v under P and let $\hat{P}(v)$ be Algorithm 5's estimate of that probability given the observed values $x_1, \ldots, x_n$. By definition,

$$\left\| \hat{P}\left(x_{1:n}\right) - P \right\|_{\infty} = \max_{v \in [d]} \left| \hat{P}(v) - P(v) \right|. \tag{108}$$

Our algorithm constructs $\hat{P}(v)$ as $\hat{P}(v) = \langle \bar{\mathbf{z}}, \mathbf{\Phi} \mathbf{e}_v \rangle$. As $\sum_i w_i = 1$, we can rewrite $P(v)$ as $\mathbb{E}_{x_{1:n}}\left[\langle \sum_i w_i \mathbf{e}_{x_i}, \mathbf{e}_v \rangle \right]$. This leads us to:

$$\begin{aligned}
\max_{v \in [d]} \left| \hat{P}(v) - P(v) \right| &= \max_{v \in [d]} \left| \langle \bar{\mathbf{z}}, \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}_{x_{1:n}}\left[\left\langle \sum_{i=1}^{n} w_i \mathbf{e}_{x_i}, \mathbf{e}_v \right\rangle \right] \right| \\
&= \max_{v \in [d]} \left| \langle \bar{\mathbf{z}} - \mathbb{E}[\bar{\mathbf{z}}] + \mathbb{E}[\bar{\mathbf{z}}], \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}_{x_{1:n}}\left[\left\langle \sum_{i=1}^{n} w_i \mathbf{e}_{x_i}, \mathbf{e}_v \right\rangle \right] \right| \\
&= \max_{v \in [d]} \left| \langle \bar{\mathbf{z}} - \mathbb{E}[\bar{\mathbf{z}}], \mathbf{\Phi} \mathbf{e}_v \rangle + \langle \mathbb{E}[\bar{\mathbf{z}}], \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}_{x_{1:n}}\left[\left\langle \sum_{i=1}^{n} w_i \mathbf{e}_{x_i}, \mathbf{e}_v \right\rangle \right] \right|.
\end{aligned}$$

We can separate these terms by applying Cauchy-Schwarz:

$$\max_{v \in [d]} \left| \hat{P}(v) - P(v) \right| \leq \max_{v \in [d]} \left| \langle \bar{\mathbf{z}} - \mathbb{E}[\bar{\mathbf{z}}], \mathbf{\Phi} \mathbf{e}_v \rangle \right| + \max_{v \in [d]} \left| \langle \mathbb{E}[\bar{\mathbf{z}}], \mathbf{\Phi} \mathbf{e}_v \rangle - \mathbb{E}_{x_{1:n}}\left[\left\langle \sum_{i=1}^{n} w_i \mathbf{e}_v, \mathbf{e}_v \right\rangle \right] \right|. \tag{109}$$

Applying Claim D.1 and a union bound, with probability at least $1 - \beta/2$, the first term of Equation (109) satisfies, for all $v \in [d]$

$$\max_{v \in [d]} \left| \langle \bar{\mathbf{z}} - \mathbb{E}[\bar{\mathbf{z}}], \mathbf{\Phi} \mathbf{e}_v \rangle \right| \leq \sqrt{ 2 \log\left(\frac{4d}{\beta} \right) \cdot \left(\sum_{i=1}^{n} \frac{1}{c_i^2} \right)^{-1} }. \tag{110}$$

To bound the second term, we can express $\mathbb{E}[\bar{\mathbf{z}}]$ as $\mathbb{E}_{x_{1:n}}[\mathbb{E}[\bar{\mathbf{z}} \mid x_{1:n}]]$, where $x_{1:n} \in [d]^n$ is the vector of values observed by the algorithm. Conditioned on knowing $x_{1:n}$, the expected value of $\bar{\mathbf{z}}$ is:

$$\mathbb{E}[\bar{\mathbf{z}} \mid x_{1:n}] = \mathbb{E}\left[\sum_{i=1}^{n} w_i z_i \mid x_{1:n} \right] = \mathbf{\Phi} \sum_{i=1}^{n} w_i \mathbf{e}_{x_i}.$$

Therefore, the second term of Equation (109) is:

$$\begin{aligned}
&\max_{v\in[d]}\left|\langle\mathbb{E}[\bar{\mathbf{z}}],\mathbf{\Phi}\mathbf{e}_v\rangle-\mathbb{E}_{x_{1:n}}\left[\left\langle\sum_{i=1}^n w_i\mathbf{e}_{x_i},\mathbf{e}_v\right\rangle\right]\right| \\
&\quad=\max_{v\in[d]}\left|\left\langle\mathbb{E}_{x_{1:n}}\left[\mathbf{\Phi}\sum_{i=1}^n w_i\mathbf{e}_{x_i}\right],\mathbf{\Phi}\mathbf{e}_v\right\rangle-\mathbb{E}_{x_{1:n}}\left[\left\langle\sum_{i=1}^n w_i\mathbf{e}_{x_i},\mathbf{e}_v\right\rangle\right]\right| \\
&\quad=\max_{v\in[d]}\left|\mathbb{E}_{x_{1:n}}\left[\left\langle\mathbf{\Phi}\sum_{i=1}^n w_i\mathbf{e}_{x_i},\mathbf{\Phi}\mathbf{e}_v\right\rangle-\left\langle\sum_{i=1}^n w_i\mathbf{e}_{x_i},\mathbf{e}_v\right\rangle\right]\right|.
\end{aligned}$$

By the Johnson-Lindenstrauss lemma (Lemma 4.1), with probability at least $1-\beta/2$,

$$\begin{aligned}
\max_{v\in[d]}\left|\mathbb{E}_{x_{1:n}}\left[\left\langle\mathbf{\Phi}\sum_{i=1}^n w_i\mathbf{e}_{x_i},\mathbf{\Phi}\mathbf{e}_v\right\rangle-\left\langle\sum_{i=1}^n w_i\mathbf{e}_{x_i},\mathbf{e}_v\right\rangle\right]\right| &\le\max_{v\in[d]}\left|\gamma\cdot\mathbb{E}_{x_{1:n}}\left[\mathcal{O}\left(\left\|\sum_{i=1}^n w_i\mathbf{e}_{x_i}\right\|^2+\|\mathbf{e}_v\|^2\right)\right]\right| \\
&=\mathcal{O}(\gamma). \qquad (111)
\end{aligned}$$

Recall that we choose

$$\gamma=\mathcal{O}\left(\sqrt{\log\left(\frac{d}{\beta}\right)\left(\sum_{i=1}^n\frac{1}{c_i^2}\right)^{-1}}\right).$$

Combining Equation (110) and Equation (111), along with this choice of γ and the fact that $\frac{1}{c_i^2}=\mathcal{O}\left(\varepsilon_i^2\right)$ for all i, the following bound on the error of our mechanism for histogram estimation holds with probability at least $1-\beta$:

$$\max_{v\in[d]}\left|\hat{P}(v)-P(v)\right|\le\mathcal{O}\left(\sqrt{\frac{\log\left(d/\beta\right)}{\sum_{i=1}^n\varepsilon_i^2}}\right). \qquad (112)$$

The upper bound of 1 on the error described in the theorem statement follows by outputting $\hat{P}(v)=0$ for all v but one v', for which we output $\hat{P}(v')=1$. □