Efficient Estimation of a Gaussian Mean with Local Differential Privacy

Nikita Kalinin
Institute of Science and Technology Austria (ISTA)

Lukas Steinberger
University of Vienna

Abstract

In this paper, we study the problem of estimating the unknown mean θ of a unit variance Gaussian distribution in a locally differentially private (LDP) way. In the high-privacy regime ($\epsilon \leq 1$), we identify an optimal privacy mechanism that minimizes the variance of the estimator asymptotically. Our main technical contribution is the maximization of the Fisher-Information of the sanitized data with respect to the local privacy mechanism Q. We find that the exact solution $Q_{\theta,\epsilon}$ of this maximization is the sign mechanism that applies randomized response to the sign of $X_i - \theta$, where $X_1, \ldots, X_n$ are the confidential iid original samples. However, since this optimal local mechanism depends on the unknown mean θ, we employ a two-stage LDP parameter estimation procedure which requires splitting agents into two groups. The first n_1 observations are used to consistently but not necessarily efficiently estimate the parameter θ by $\tilde{\theta}_{n_1}$. Then this estimate is updated by applying the sign mechanism with $\tilde{\theta}_{n_1}$ instead of θ to the remaining $n - n_1$ observations, to obtain an LDP and efficient estimator of the unknown mean.

1 Introduction

We consider the problem of estimating the unknown mean of a unit variance normal distribution in a locally differentially private and statistically efficient way. In this scenario, we have n agents or data-holders, each of which owns data X_i sampled independently from the probability distribution $P_\theta = N(\theta, 1)$ which depends on the unknown mean parameter $\theta \in \mathbb{R}$. Denote the class of all those potential data generating distributions by $\mathcal{P} = \{P_\theta : \theta \in \mathbb{R}\}$. We protect the privacy of the data owners by local differential

privacy. This means that each agent generates a sanitized version Z_i of their original data X_i independently of everybody else by applying a privacy mechanism Q, which is a Markov kernel or a conditional distribution of Z_i given $X_i = x$. In other words, $Q(A|x) = P(Z_i \in A | X_i = x)$. From the privacy mechanism Q we can generate random variates in some arbitrary space $\mathcal{Z}$ endowed with a sigma algebra $\mathcal{G}$. We say that Q satisfies the ϵ-local differential privacy property if

$$Q(A|x) \leq e^\epsilon Q(A|x') \tag{1}$$

for all $x, x' \in \mathbb{R}$ and all events $A \in \mathcal{G}$. We denote by $\mathcal{Q}_\epsilon = \mathcal{Q}_\epsilon(\mathbb{R})$ the set of all possible such privacy mechanisms taking inputs from $\mathbb{R}$.

After each agent applied the privacy mechanism Q we obtain iid samples $Z_1, \ldots Z_n$ from the distribution QP_θ, which formally is the probability measure obtained by integrating Q with respect to $P_\theta(dx)$. We can only use these samples to estimate the unknown parameter θ and our goal is to do this with the smallest amount of estimation variance possible, that is, to estimate θ in a statistically efficient way. Standard asymptotic theory (cf. van der Vaart, 2007, Chapter 8) shows that a maximum likelihood estimator $\hat{\theta}_n^{(mle)}$ based on $Z_1, \ldots, Z_n$ achieves an asymptotic normal distribution

$$\sqrt{n}(\hat{\theta}_n^{(mle)} - \theta) \xrightarrow[n \to \infty]{d} N(0, I_\theta(Q\mathcal{P})^{-1}) \tag{2}$$

where the inverse of the Fisher-Information

$$I_\theta(Q\mathcal{P}) := \mathbb{E}_{Z \sim QP_\theta}\left[\left(\frac{\partial \log QP_\theta(Z)}{\partial \theta}\right)^2\right] \tag{3}$$

is the smallest variance achievable by regular estimators (van der Vaart, 2007, Theorem 8.8). Thus, to minimize asymptotic variance, we should solve the optimization problem

$$\max_Q I_\theta(Q\mathcal{P}) \quad \text{subject to} \quad Q \in \mathcal{Q}_\epsilon. \tag{4}$$

This is a very challenging optimization problem because the set of Markov kernels $\mathcal{Q}_\epsilon$ is a huge, non-parametric set and we have to even search over all possible output spaces $(\mathcal{Z}, \mathcal{G})$ from which Q can draw Z_i. Furthermore,

it turns out that for fixed output space $(\mathcal{Z}, \mathcal{G})$ the mapping $Q \mapsto I_\theta(Q\mathcal{P})$ is convex, but we are trying to maximize this objective function. Hence, we may be confronted with numerous local and global optima at the boundary of the feasible set $\mathcal{Q}_\epsilon$. Steinberger (2024) provided a numerical scheme for approximately solving (4) based on a linear program representation of Kairouz et al. (2016). However, the runtime of the linear program scales exponentially in the accuracy of approximation. Steinberger (2024) also showed that indeed, $[\sup_{Q \in \mathcal{Q}_\epsilon} I_\theta(Q\mathcal{P})]^{-1}$ is the smallest possible asymptotic variance among all (sequentially interactive) locally differentially private and regular estimators. Here, regular means that the scaled estimation error $\sqrt{n}(\hat{\theta}_n - \theta_n)$ converges in distribution to a limiting distribution along the sequence of true parameters $\theta_n = \theta + h/\sqrt{n}$ for all $h \in \mathbb{R}$ and such that the limiting distribution does not depend on h (see van der Vaart (2007, Section 8.5) and Theorem 2 below). Regularity is indeed necessary to establish the mentioned lower bound on the asymptotic variance. Without regularity one could still show that the lower bound holds at least for all $\theta \in \mathbb{R}$ outside of a Lebesgue null set (see van der Vaart (2007, Theorem 8.8) for the classical case and the discussion at the end of Section 3 in Steinberger (2024) for the locally private case), but certain pathologies may arise for irregular estimators (see Lehmann and Casella, 1998, Example 2.5 and Example 2.7). Fortunately, we can show in Theorem 2 below that our proposed locally private estimator is also regular and achieves the asymptotic lower bound.

Our main contribution is to show that the following simple **sign mechanism** is an exact solution of (4) in the case where the privacy parameter ϵ is sufficiently small: Given that $X_i = x$, we generate Z_i by

$$Z_i = \begin{cases} \text{sgn}(x - \theta) & \text{with probability } p_\epsilon \\ -\text{sgn}(x - \theta) & \text{with probability } 1 - p_\epsilon, \end{cases} \quad (5)$$

where $p_\epsilon := \frac{e^\epsilon}{1 + e^\epsilon}$ and $\text{sgn}(0) := 1$. In other words, for $z \in \mathcal{Z} = \{-1, 1\}$ and $x \in \mathbb{R}$, we set

$$Q_{\theta, \epsilon}^{\text{sgn}}(\{z\}|x) := P(Z_i = z | X_i = x)$$
$$= \begin{cases} p_\epsilon, & \text{if } z = \text{sgn}(x - \theta), \\ 1 - p_\epsilon, & \text{if } z \neq \text{sgn}(x - \theta). \end{cases} \quad (6)$$

Notice that this mechanism can also be represented as a composition $Q_{\theta, \epsilon}^{\text{sgn}} = Q_\epsilon^{RR} T_\theta$, where $T_\theta(x) := \text{sgn}(x - \theta)$ and Q_ϵ^{RR} is the randomized response mechanism of Warner (1965) that flips the sign of the input with probability $1 - p_\epsilon$. In particular, we easily conclude that $Q_{\theta, \epsilon}^{\text{sgn}} \in \mathcal{Q}_\epsilon$. We can now state and prove the following result.

Theorem 1. *If $\epsilon \leq 1.04$ then*

$$I_\theta(Q\mathcal{P}) \leq I_\theta(Q_{\theta, \epsilon}^{\text{sgn}} \mathcal{P}) = \frac{2}{\pi} \frac{(e^\epsilon - 1)^2}{(e^\epsilon + 1)^2}, \quad (7)$$

for all $\theta \in \mathbb{R}$ and all $Q \in \mathcal{Q}_\epsilon$. In particular, the sign mechanism solves (4).

Our proof is based on the idea of quantizing the normal distribution and solving the problem of maximizing the private Fisher Information for discrete distributions simultaneously for all quantization levels. We apply results from Kairouz et al. (2016) and a delicate duality argument to solve the discrete case. The main steps of the proof are collected in Subsection 2.1.

An obvious issue with the sign mechanism in (6) is that it depends on the unknown parameter θ which we want to estimate. By incorporating the sign mechanism into a two-stage scheme as in Steinberger (2024) we end up with the following procedure for asymptotically efficient parameter estimation: Split the agents into two groups of size n_1 and $n_2 := n - n_1$ such that $\lim \frac{n_1}{n} = 0$ and $n_1(n) \to \infty$.[1] Use the first group to get a consistent but not necessarily efficient estimate of the unknown mean θ and the second group to update the first-stage estimate to make it asymptotically efficient. The two stages of the procedure are described below, where θ_0 is an arbitrary initial guess for the unknown mean θ.

1. • Apply the sign mechanism $Q_{\theta_0, \epsilon}^{\text{sgn}}$ at the initial value θ_0 to $X_1, \dots, X_{n_1}$ to obtain iid sanitized data $Z_i \sim Q_{\theta_0, \epsilon}^{\text{sgn}} P_\theta, i = 1, \dots, n_1$.
   • Compute

$$\tilde{\theta}_{n_1} = \theta_0 - \Phi^{-1}\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon + 1}{e^\epsilon - 1}\bar{Z}_{n_1}\right), \quad (8)$$

   if $|\bar{Z}_{n_1}| < \frac{e^\epsilon - 1}{e^\epsilon + 1}$, otherwise $\tilde{\theta}_{n_1} = \theta_0$, where Φ is the cumulative distribution function of the standard normal distribution and $\bar{Z}_{n_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} Z_i$.

2. • Apply the sign mechanism $Q_{\tilde{\theta}_{n_1}, \epsilon}^{\text{sgn}}$ at the first stage estimate $\tilde{\theta}_{n_1}$ to the data $X_{n_1+1}, \dots, X_n$ in the second group to generate values $Z_i \sim Q_{\tilde{\theta}_{n_1}, \epsilon}^{\text{sgn}} P_\theta, i = n_1 + 1, \dots, n$.
   • Compute

$$\hat{\theta}_n = \tilde{\theta}_{n_1} - \Phi^{-1}\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon + 1}{e^\epsilon - 1}\bar{Z}_{n_2}\right), \quad (9)$$

   if $|\bar{Z}_{n_2}| < \frac{e^\epsilon - 1}{e^\epsilon + 1}$, otherwise $\hat{\theta}_n = \tilde{\theta}_{n_1}$ where $\bar{Z}_{n_2} = \frac{1}{n_2}\sum_{i=n_1+1}^{n} Z_i$.

Relying on our main Theorem 1, we show that the above two-stage estimation procedure is regular and asymptotically achieves minimal variance. Let R_θ denote the distribution of the full sanitized data $Z_1, \dots, Z_n$ when the

---

[1] We parametrize $n_1 = n_1(n)$ by n such that all limits are understood as $n \to \infty$.

true unknown mean of the X_i is $\theta \in \mathbb{R}$. Thus, $R_{\theta+h/\sqrt{n}}$ is the distribution of the sanitized data when the true unknown mean of the X_i is $\theta_n := \theta + h/\sqrt{n}$. Hence, on top of asymptotic normality with optimal asymptotic variance (i.e., the case $h = 0$), the following result also establishes regularity of the locally private estimator sequence $\hat{\theta}_n$. Please refer to Appendix A.1 for the proof.

Theorem 2. *If $\epsilon \leq 1.04$ then the two-stage locally private estimation procedure described above satisfies*

$$\sqrt{n}\left(\hat{\theta}_n - \theta - \frac{h}{\sqrt{n}}\right) \xrightarrow[n\to\infty]{R_{\theta+\frac{h}{\sqrt{n}}}} N\left(0, \left[\sup_{Q \in \mathcal{Q}_\epsilon} I_\theta(Q\mathcal{P})\right]^{-1}\right),$$

for any value θ_0, h and $n_1 = o(n)$ such that $n_1(n) \to \infty$.

Notice, however, that the two-stage estimator $\hat{\theta}_n$ for a finite n depends on the choice of initial value θ_0 and on the size n_1 of the first group. Hence, θ_0 and n_1 are tuning parameters in our estimation procedure. In Section 3, we investigate the impact of θ_0 and n_1 on the performance of our locally private estimator in a finite sample simulation study.

1.1 Related Literature

Local differential privacy originated about 20 years ago (see Dinur and Nissim, 2003; Dwork and Nissim, 2004; Dwork et al., 2006; Dwork, 2008; Evfimievski et al., 2003) and has since become an incredibly popular way of data privacy protection when there is no trusted third party available. Despite its popularity, some of the most basic learning problems, such as the one considered here, have not been fully solved under LDP. Much attention has recently been paid to discrete distribution estimation and deriving upper bounds on the estimation error. For instance, Wang et al. (2016) have suggested an optimal mechanism for mutual information maximization for discrete distributions under LDP. Ye and Barg (2018) are studying the locally private minimax risk for discrete distributions. Nam and Lee (2022) is perhaps closer to our work. They consider the problem of maximizing a Fisher-Information in the local privacy mechanism, but they restrict to the 1-bit communication constraint which, for our purpose, is an oversimplified case. Barnes et al. (2020) also investigate and bound Fisher-Information when original data are perturbed using a differentially private randomization mechanism. But they have not attempted to find an optimal randomization mechanism or an efficient estimation procedure. Gaussian mean estimation is considered by Joseph et al. (2019) and they also use a two-stage procedure for parameter estimation. They obtain high-probability bounds for the estimation error, which is of order $n^{-1/2}$ for the two-stage protocol, but unlike our work, they do not obtain the optimal asymptotic constant. Duchi and Rogers (2019) proposed a one-stage algorithm for Gaussian mean estimation, which resembles

our first-stage estimator. While their method can outperform a two-stage protocol, this advantage holds primarily when the initial guess θ_0 is close to the true parameter θ and the sample size is relatively small. Asi et al. (2022) introduced the PrivUnitG algorithm, designed specifically for high-dimensional problems where the data resides on the unit sphere. While their method is effective in these settings, it is worth noting that for one-dimensional data, where the values are ± 1, their approach differs from the randomized response, which is provably optimal for this problem. Consequently, their method is inefficient for one-dimensional cases. Building on these insights, Asi et al. (2023) developed the ProjUnit framework, which further optimizes mean estimation by projecting data into lower-dimensional subspaces before applying an optimal randomizer. A recent work by Liu et al. (2023) introduced a continual, gradient-like procedure to iteratively refine the estimate of θ. They have shown that the procedure achieves a variance that is optimal among binary mechanisms. However, we will show that this variance is asymptotically optimal among all LDP mechanisms.

The conceptual foundation to the problem of finding the optimal Fisher-Information mechanism was laid by Steinberger (2024). However, while in that work the optimal mechanism and locally private estimation procedure have to be approximated by a computationally expensive numerical optimization routine, we here derive an exact and simple closed-form privacy mechanism and estimator.

2 Proof of main results

2.1 Proof of Theorem 1

In this section, we prove Theorem 1. We first follow the approach of Steinberger (2024) and quantize the Fisher-Information to then solve a discrete version of (4). However, while Steinberger (2024) had to rely on a numerical algorithm for the discrete optimization, which has exponential runtime in the resolution of the discretization, we here solve all the discrete problems exactly by providing a closed form solution and show that it is the same for all resolution levels k of the approximation. We then conclude that the solution to the simplified discrete problems is also the global solution. Notice that all the regularity conditions imposed by Steinberger (2024) are satisfied for the Gaussian location model $\mathcal{P} = \{N(\theta, 1) : \theta \in \mathbb{R}\}$ that we consider here (cf. Section 5.2 in Steinberger, 2024). In particular, the Fisher-Information $I_\theta(Q\mathcal{P})$ is well defined and finite for any privacy mechanism $Q \in \mathcal{Q}_\epsilon$.

2.1.1 Discrete approximation

We begin by defining what is called a consistent quantizer in Steinberger (2024, cf. Definition 3 and Section 5 in that reference). For an even positive integer k and for $j =$

$1, \ldots, k - 1$, let $B_j := (x_{j-1}, x_j]$ and $B_k := (x_{k-1}, \infty)$, where $x_j := \Phi^{-1}(j/k)$ and Φ is the cdf of the standard normal distribution. Now set $T_{k,\theta}(x) := \sum_{j=1}^{k} j \mathbb{1}_{B_j}(x - \theta)$. Notice that $T_{k,\theta}$ maps the real line $\mathbb{R}$ into the discrete set $[k] := \{1, \ldots, k\}$, hence, it is called a quantizer. We write $T_{k,\theta_0}\mathcal{P} := \{r_{\theta,\theta_0} : \theta \in \mathbb{R}\}$ for the resulting quantized model, where $r_{\theta,\theta_0}(j) := P_\theta(T_{k,\theta_0}^{-1}(\{j\})) = P_\theta(B_j + \theta_0) = \Phi(x_j + \theta_0 - \theta) - \Phi(x_{j-1} + \theta_0 - \theta)$ is the probability mass function of the quantized data $T_{k,\theta_0}(X_i)$ and we set $\Phi(-\infty) := 0$, $\Phi(\infty) := 1$, $\phi(x) := \Phi'(x)$ and $\phi(\pm\infty) := 0$. Lemmas 4.7, 4.9, 5.1 and 5.2 in Steinberger (2024) show that for every even k and every $\theta \in \mathbb{R}$

$$\sup_{Q \in \mathcal{Q}_\epsilon(\mathbb{R})} I_\theta(Q\mathcal{P}) \leq \sup_{Q \in \mathcal{Q}_\epsilon([k])} I_\theta(QT_{k,\theta}\mathcal{P}) + \Delta_k \quad (10)$$

$$= \sup_{Q \in \mathcal{Q}_\epsilon([k] \to [k])} I_\theta(QT_{k,\theta}\mathcal{P}) + \Delta_k, \quad (11)$$

where $\Delta_k \to 0$ as $k \to \infty$, $\mathcal{Q}_\epsilon([k])$ is the set of ϵ-private mechanisms that take inputs from $[k]$ and produce random outputs from some arbitrary measurable space $(\mathcal{Z}, \mathcal{G})$, and $\mathcal{Q}_\epsilon([k] \to [k])$ are the ϵ-private mechanisms whose inputs and outputs both take values in $[k]$. Thus, the elements of $\mathcal{Q}_\epsilon([k] \to [k])$ can be represented as $k \times k$ column stochastic matrices with the property that $Q_{ij} \leq e^\epsilon Q_{ij'}$, for all $i, j, j' \in [k]$. Since obviously $\sup_{Q \in \mathcal{Q}_\epsilon([k] \to [k])} I_\theta(QT_{k,\theta}\mathcal{P}) \leq \sup_{Q \in \mathcal{Q}_\epsilon} I_\theta(Q\mathcal{P})$, (10) yields

$$\sup_{Q \in \mathcal{Q}_\epsilon([k] \to [k])} I_\theta(QT_{k,\theta}\mathcal{P}) \to \sup_{Q \in \mathcal{Q}_\epsilon} I_\theta(Q\mathcal{P}), \quad \forall \theta \in \mathbb{R},$$

as $k \to \infty$, k even. In the remainder of the proof, we will show that $\sup_{Q \in \mathcal{Q}_\epsilon([k] \to [k])} I_\theta(QT_{k,\theta}\mathcal{P}) = I_\theta(Q_{\theta,\epsilon}^{\mathrm{sgn}}\mathcal{P})$ for every even k. Thus, the sign mechanism in (6) must be a solution of (4).

2.1.2 Evaluating the quantized objective function

Next, we provide an explicit expression for the quantized private Fisher-Information $I_\theta(QT_{k,\theta_0}\mathcal{P})$, for arbitrary $Q \in \mathcal{Q}_\epsilon([k] \to [k])$. The quantized and Q-privatized model $QT_{k,\theta_0}\mathcal{P} = \{m_{\theta,\theta_0} : \theta \in \mathbb{R}\}$ can be expressed via its probability mass functions $m_{\theta,\theta_0}(i) := \sum_{j=1}^{k} Q_{ij} r_{\theta,\theta_0}(j)$, $i \in [k]$. Thus, we have

$$I_\theta(QT_{k,\theta_0}\mathcal{P}) = \mathbb{E}_{Z \sim QT_{k,\theta_0}P_\theta} \left[\left(\frac{\partial \log m_{\theta,\theta_0}(Z)}{\partial \theta} \right)^2 \right]$$

$$= \sum_{i=1}^{k} \frac{\dot{m}_{\theta,\theta_0}(i)^2}{m_{\theta,\theta_0}(i)},$$

where $\dot{m}_{\theta,\theta_0}(i) := \frac{\partial}{\partial \theta} m_{\theta,\theta_0}(i) = \sum_{j=1}^{k} Q_{ij}[\phi(x_{j-1} + \theta_0 - \theta) - \phi(x_j + \theta_0 - \theta)]$ and the ratio is to be understood as equal to zero if the denominator $m_{\theta,\theta_0}(i)$ is zero.

Abbreviating $y_j := \phi(x_{j-1}) - \phi(x_j)$, we arrive at

$$I_\theta(QT_{k,\theta}\mathcal{P}) = \sum_{i=1}^{k} \frac{\left(\sum_{j=1}^{k} Q_{ij} y_j\right)^2}{\frac{1}{k} \sum_{j=1}^{k} Q_{ij}} = \sum_{i=1}^{k} \mu(Q_{i\cdot}^T),$$

$$(12)$$

for $\mu(v) := k \frac{(v^T y)^2}{v^T \vec{1}}$, $v \in \mathcal{C}_k := \{u \in \mathbb{R}_+^k : u_j \leq e^\epsilon u_{j'}, \forall j, j' \in [k]\}$, and $\mu(0) := 0$. In particular, we see that $I_\theta(QT_{k,\theta}\mathcal{P}) = I_0(QT_{k,0}\mathcal{P})$ for all $\theta \in \mathbb{R}$. For later use we note that $y_j = -y_{k-j+1}$, $\sum_{j=1}^{k/2} y_j = -\phi(0) = -\sum_{j=k/2+1}^{k} y_j$ and $\sum_{j=1}^{k} y_j = 0$.

We conclude this subsection by computing the Fisher-Information for the sign mechanism in (6). Use the fact that $Q_{\theta,\epsilon}^{\mathrm{sgn}} = Q_\epsilon^{RR} \circ T_{2,\theta} \in \mathcal{Q}_\epsilon([2] \to [2])$, where the randomized response mechanism can be represented in matrix form as

$$Q_\epsilon^{RR} = \begin{pmatrix} \frac{e^\epsilon}{1+e^\epsilon} & \frac{1}{1+e^\epsilon} \\ \frac{1}{1+e^\epsilon} & \frac{e^\epsilon}{1+e^\epsilon} \end{pmatrix}. \quad (13)$$

Thus,

$$I_\theta(Q_{\theta,\epsilon}^{\mathrm{sgn}}\mathcal{P}) = I_\theta(Q_\epsilon^{RR} T_{2,\theta}\mathcal{P}) = \sum_{i=1}^{2} \frac{\left(\sum_{j=1}^{2} [Q_\epsilon^{RR}]_{ij} y_j\right)^2}{\frac{1}{2} \sum_{j=1}^{2} [Q_\epsilon^{RR}]_{ij}}$$

$$= \frac{1}{1+e^\epsilon} \left(\frac{\phi(0)^2(1-e^\epsilon)^2}{\frac{1}{2}(1+e^\epsilon)} + \frac{\phi(0)^2(e^\epsilon-1)^2}{\frac{1}{2}(1+e^\epsilon)} \right)$$

$$= \frac{2}{\pi} \frac{(e^\epsilon-1)^2}{(e^\epsilon+1)^2}.$$

2.1.3 Reformulation as a Linear Program

From Lemma 4.5 of Steinberger (2024), $Q \mapsto I_0(QT_{k,0}\mathcal{P})$ is a continuous, sublinear and convex function on the compact set $\mathcal{Q}_\epsilon([k] \to [k]) \subseteq \mathbb{R}^{k \times k}$. Therefore, a maximizer exists and Theorems 2 and 4 of Kairouz et al. (2016) allow us to conclude that the maximization problem

$$\max_Q I_0(QT_{k,0}\mathcal{P}) \quad \text{subject to} \quad Q \in \mathcal{Q}_\epsilon([k] \to [k]) \quad (14)$$

has the same optimal value as the linear program

$$\max_{\alpha \in \mathbb{R}^{2^k}} \quad \sum_{j=1}^{2^k} \mu(S_{\cdot j}^{(k)}) \alpha_j$$

$$\text{s.t.} \quad S^{(k)} \alpha = \vec{1}$$

$$\alpha \geq 0, \quad (15)$$

where $S^{(k)}$ is a staircase matrix defined as follows: For $0 \leq j \leq 2^k - 1$ consider its binary representation $b_j \in \{0,1\}^k$ with $j = \sum_{i=1}^{k} b_{ij} 2^{k-i}$. Then $(S^{(k)})_{i,j+1} := b_{ij}(e^\epsilon - 1) + 1$. For instance, with $k = 3$ we get the following matrix:

$$S^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 1 & e^\epsilon & e^\epsilon & e^\epsilon & e^\epsilon \\ 1 & 1 & e^\epsilon & e^\epsilon & 1 & 1 & e^\epsilon & e^\epsilon \\ 1 & e^\epsilon & 1 & e^\epsilon & 1 & e^\epsilon & 1 & e^\epsilon \end{bmatrix}.$$

Moreover, if α^* is a solution of (15), then $Q^* = [S^{(k)}\mathrm{diag}(\alpha^*)]^T \in \mathbb{R}^{2^k \times k}$ is a solution of (14). From Kairouz et al. (2016, Theorem 2) we know that an optimal mechanism Q^* has at most k non-zero rows. Since zero rows do not contribute to the Fisher-Information (12), we can remove them from Q^* to obtain $Q^* \in \mathbb{R}^{k \times k}$.

Finally, notice that $\alpha_j^* := (1 + e^\epsilon)^{-1}$ for $j \in \{2^{k/2}, 2^k - 2^{k/2} + 1\}$ and $\alpha_j^* := 0$ else, is a feasible point of (15), because $b_{2^{k/2}} = \vec{1} - b_{2^k - 2^{k/2}+1}$ and thus

$$
S^{(k)}\alpha^* = \frac{1}{1 + e^\epsilon}\left[(b_{2^{k/2}} + b_{2^k - 2^{k/2}+1})(e^\epsilon - 1) + 2 \cdot \vec{1}\right]
$$
$$
= \vec{1}.
$$

Furthermore, α^* achieves an objective function value of

$$
\frac{1}{1 + e^\epsilon}\left[\mu(b_{2^{k/2}}(e^\epsilon - 1) + \vec{1}) + \mu(b_{2^k - 2^{k/2}+1}(e^\epsilon - 1) + \vec{1})\right]
$$
$$
= \frac{2}{1 + e^\epsilon}\left[\frac{\phi(0)^2(e^\epsilon - 1)^2}{e^\epsilon + 1} + \frac{\phi(0)^2(e^\epsilon - 1)^2}{e^\epsilon + 1}\right]
$$
$$
= I_\theta(Q_{\theta,\epsilon}^{\mathrm{sgn}}\mathcal{P}).
$$

Hence, we have $\sup_{Q \in \mathcal{Q}_\epsilon([k] \to [k])} I_\theta(QT_{k,\theta}\mathcal{P}) = \sup_{Q \in \mathcal{Q}_\epsilon([k] \to [k])} I_0(QT_{k,0}\mathcal{P}) \geq I_\theta(Q_{\theta,\epsilon}^{\mathrm{sgn}}\mathcal{P})$, for every even $k > 0$. It remains to establish an upper bound, which we do by a duality argument.

2.1.4 Guessing a dual solution

In this section, we study the dual of the LP in (15). By weak duality, any feasible value of the dual provides an upper bound on the objective function of the primal problem. Hence, the challenge is to identify a feasible value of the dual that achieves an objective function equal to $I_\theta(Q_{\theta,\epsilon}^{\mathrm{sgn}}\mathcal{P})$. The dual program reads as follows:

$$
\min_{\beta \in \mathbb{R}^k} \quad \vec{1}^T\beta
$$
$$
\text{s.t.} \quad (S^{(k)})^T\beta \geq \vec{\mu} \tag{16}
$$

where $\vec{\mu} := (\mu_j)_{j=1}^{2^k} := (\mu(S_{\cdot j}^{(k)}))_{j=1}^{2^k} \in \mathbb{R}^{2^k}$. The proof of Theorem 1 is finished if we can identify a feasible point $\beta \in \mathbb{R}^k$ of (16) satisfying $\vec{1}^T\beta = \sum_{j=1}^k \beta_j = \frac{2}{\pi}t_\epsilon^2$, where $t_\epsilon := \frac{e^\epsilon - 1}{e^\epsilon + 1}$. Our guess for such a $\beta^* \in \mathbb{R}^k$ is

$$
\beta_j^* := -\frac{2t_\epsilon^2}{\pi k} + |y_j|t_\epsilon^2\sqrt{\frac{8}{\pi}}. \tag{17}
$$

This clearly satisfies $\vec{1}^T\beta^* = \frac{2}{\pi}t_\epsilon^2$. It is the main technical challenge of the proof to show that β^* is a feasible point of (16).

For fixed $j \in [2^k]$, we have to verify $(S_{\cdot j}^{(k)})^T\beta^* \geq \mu(S_{\cdot j}^{(k)})$. It will be convenient to partition the index set $[k]$ as follows $[k] = I_1^0 \cup I_{e^\epsilon}^0 \cup I_1^1 \cup I_{e^\epsilon}^1$ where $I_1^0 = \{i | i \leq \frac{k}{2}, S_{i,j}^{(k)} = 1\}$,

$I_{e^\epsilon}^0 = \{i| i \leq \frac{k}{2}, S_{i,j}^{(k)} = e^\epsilon\}$ and $I_1^1, I_{e^\epsilon}^1$ defined similarly for the second half of indices $i = \frac{k}{2} + 1, \ldots, k$. Then we can rewrite the required inequality $(S_{\cdot j}^{(k)})^T\beta^* \geq \mu(S_{\cdot j}^{(k)})$ as follows:

$$
\sum_{i \in I_1^0}\beta_i^* + e^\epsilon\sum_{i \in I_{e^\epsilon}^0}\beta_i^* + \sum_{i \in I_1^1}\beta_i^* + e^\epsilon\sum_{i \in I_{e^\epsilon}^1}\beta_i^*
$$
$$
\geq \frac{\left(\sum_{i \in I_1^0}y_i + e^\epsilon\sum_{i \in I_{e^\epsilon}^0}y_i + \sum_{i \in I_1^1}y_i + e^\epsilon\sum_{i \in I_{e^\epsilon}^1}y_i\right)^2}{\frac{1}{k}(|I_1^0| + |I_1^1| + e^\epsilon|I_{e^\epsilon}^0| + e^\epsilon|I_{e^\epsilon}^1|)}.
$$

For convenience let us denote $m_1 = |I_1^0|$, $m_2 = |I_{e^\epsilon}^1|$, $a_1 = \sqrt{2\pi}\sum_{i \in I_1^0}|y_i|$, $a_2 = \sqrt{2\pi}\sum_{i \in I_{e^\epsilon}^1}|y_i|$ such that $m_1, m_2 \leq \frac{k}{2}$ and $a_1, a_2 \in [0, 1]$. With this new notation, we can rewrite the sum on the left-hand-side as

$$
(1 - e^\epsilon)\sum_{i \in I_1^0}\beta_i^* + (e^\epsilon - 1)\sum_{i \in I_{e^\epsilon}^0}\beta_i^* + e^\epsilon\sum_{i=1}^{k/2}\beta_i^* + \sum_{i=k/2+1}^k\beta_i^* \tag{18}
$$

and the ratio on the right-hand-side as

$$
\frac{\left(\frac{e^\epsilon - 1}{\sqrt{2\pi}}a_1 + \frac{e^\epsilon - 1}{\sqrt{2\pi}}a_2 - \frac{e^\epsilon}{\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}}\right)^2}{\frac{1}{k}(k/2 + m_1 - m_2 + e^\epsilon(k/2 + m_2 - m_1))}. \tag{19}
$$

Substituting the values for β^* we simplify the expression in (18) to obtain

$$
(1 - e^\epsilon)\sum_{i \in I_1^0}\beta_i^* + (e^\epsilon - 1)\sum_{i \in I_{e^\epsilon}^0}\beta_i^* + e^\epsilon\sum_{i=1}^{k/2}\beta_i^* + \sum_{i=k/2+1}^k\beta_i^*
$$
$$
= (1 - e^\epsilon)\frac{2t_\epsilon^2}{\pi}\left(-\frac{m_1}{k} + a_1\right)
$$
$$
+ (e^\epsilon - 1)\frac{2t_\epsilon^2}{\pi}\left(-\frac{m_2}{k} + a_2\right) + (1 + e^\epsilon)\frac{t_\epsilon^2}{\pi}
$$
$$
= (e^\epsilon - 1)\frac{2t_\epsilon^2}{\pi}\left[\frac{m_1 - m_2}{k} + a_2 - a_1\right] + (1 + e^\epsilon)\frac{t_\epsilon^2}{\pi}.
$$

Next, we simplify (19) to get

$$
\frac{\left(\frac{e^\epsilon - 1}{\sqrt{2\pi}}a_1 + \frac{e^\epsilon - 1}{\sqrt{2\pi}}a_2 - \frac{e^\epsilon}{\sqrt{2\pi}} + \frac{1}{\sqrt{2\pi}}\right)^2}{\frac{1}{k}\left(\frac{k}{2} + m_1 - m_2 + e^\epsilon(\frac{k}{2} + m_2 - m_1)\right)}
$$
$$
= \frac{(a_1 + a_2 - 1)^2(e^\epsilon - 1)^2}{\frac{2\pi}{k}\left(\frac{k}{2}(1 + e^\epsilon) + (m_2 - m_1)(e^\epsilon - 1)\right)}
$$
$$
= \frac{(a_1 + a_2 - 1)^2(e^\epsilon - 1)^2}{\pi(1 + e^\epsilon)\left(1 + 2t_\epsilon\frac{m_2 - m_1}{k}\right)}.
$$

After these simplifications, we see that the inequality we need to verify is given by

$$(e^\epsilon - 1)\frac{2t_\epsilon^2}{\pi}\left[\frac{m_1 - m_2}{k} + a_2 - a_1\right] + (1 + e^\epsilon)\frac{t_\epsilon^2}{\pi}$$

$$\geq \frac{(a_1 + a_2 - 1)^2(e^\epsilon - 1)^2}{\pi(1 + e^\epsilon)\left(1 + 2t_\epsilon\frac{m_2 - m_1}{k}\right)}.$$

Dividing by $(1 + e^\epsilon)t_\epsilon^2/\pi$, this can be further simplified to

$$\mathcal{L}\left(a_1, a_2, \frac{m_1}{k}, \frac{m_2}{k}\right) := (a_2 - a_1)\left[2t_\epsilon + 4t_\epsilon^2\frac{m_2 - m_1}{k}\right]$$

$$- 4t_\epsilon^2\left(\frac{m_2 - m_1}{k}\right)^2 - (a_1 + a_2 - 1)^2 + 1 \geq 0.$$

(20)

Without loss of generality, we can assume $a_1 + a_2 \leq 1$; otherwise, we could have chosen $a_1 = \sqrt{2\pi}\sum_{i\in I_1^1}|y_i|$ and $a_2 = \sqrt{2\pi}\sum_{i\in I_{e^\epsilon}^0}|y_i|$, maintaining the same inequality due to the symmetry of $\beta_i^* = \beta_{k+1-i}^*$ and antisymmetry in $y_i = -y_{k+1-i}$, but with the sum $a_1 + a_2 \leq 1$, because the total sum equals $\sqrt{2\pi}\sum_{i=1}^{k}|y_i| = 2$. Given this, we can prove the inequality $\mathcal{L}\left(a_1, a_2, \frac{m_1}{k}, \frac{m_2}{k}\right) \geq 0$ using the following four lemmas.

Lemma 1. *For any k, the sequence $y_j = \phi(\Phi^{-1}((j-1)/k)) - \phi(\Phi^{-1}(j/k))$, $j = 1, \ldots, k$, is increasing.*

Proof. For $x \in (1, k]$, define $g(x) := \phi(\Phi^{-1}((x-1)/k)) - \phi(\Phi^{-1}(x/k))$ and $g(1) := -\phi(\Phi^{-1}(1/k))$. Since g is continuous on $[1, k]$, it suffices to show that g is strictly increasing on $(1, k)$. Using the fact that $\phi'(x) = (-x)\phi(x)$ and the inverse function theorem we get

$$g'(x) = \frac{1}{k}\left(\frac{\phi'}{\phi} \circ \Phi^{-1}((x-1)/k) - \frac{\phi'}{\phi} \circ \Phi^{-1}(x/k)\right)$$

$$= \frac{1}{k}\left(\Phi^{-1}(x/k) - \Phi^{-1}((x-1)/k)\right) > 0.$$

Thus, g is strictly increasing on $(1, k)$. $\square$

Lemma 2. *For $x \in [0, \frac{1}{2}]$ we have*

$$\phi(0) - \phi\left(\Phi^{-1}\left(\frac{1}{2} + x\right)\right) \geq \sqrt{\frac{\pi}{2}}x^2. \quad (21)$$

Proof. For $x \in [0, \frac{1}{2}]$, define $g(x) := \phi(0) - \phi\left(\Phi^{-1}\left(\frac{1}{2} + x\right)\right) - \sqrt{\frac{\pi}{2}}x^2$ then $g(0) = 0, g(\frac{1}{2}) = +\infty$. Since g is continuous on $[0, \frac{1}{2}]$, it suffices to show that g is strictly increasing on $(0, \frac{1}{2})$. Using the fact that $\phi'(x) = (-x)\phi(x)$ and the inverse function theorem we get

$$g'(x) = -\frac{\phi'}{\phi} \circ \Phi^{-1}\left(\frac{1}{2} + x\right) - \sqrt{2\pi}x$$

$$= \Phi^{-1}\left(\frac{1}{2} + x\right) - \sqrt{2\pi}x,$$

$$g''(x) = \frac{1}{\phi(\Phi^{-1}\left(\frac{1}{2} + x\right))} - \sqrt{2\pi} \geq 0.$$

Since the first derivative is equal to zero at zero and non-decreasing, we see that the function g itself must be non-decreasing. $\square$

By combining these two lemmas, we can easily show

$$a_2 = \sqrt{2\pi}\sum_{i\in I_{e^\epsilon}^1}y_i \geq \sqrt{2\pi}\sum_{i=k/2+1}^{k/2+m_2}y_i$$

$$= \sqrt{2\pi}\left(\phi(0) - \phi\left(\Phi^{-1}\left(\frac{1}{2} + \frac{m_2}{k}\right)\right)\right) \geq \pi\frac{m_2^2}{k^2}.$$

A similar lower bound holds for a_1 as well, namely $a_1 \geq \pi\frac{m_1^2}{k^2}$. Therefore, it is sufficient to prove the inequality (20) under restrictions that $a_1 + a_2 \leq 1$, $a_1 \geq \pi\frac{m_1^2}{k^2}$, $a_2 \geq \pi\frac{m_2^2}{k^2}$, and $m_1, m_2 \leq \frac{k}{2}$. Consider the derivative of $\mathcal{L}\left(a_1, a_2, \frac{m_1}{k}, \frac{m_2}{k}\right)$ with respect to a_2,

$$2t_\epsilon\left(1 + 2t_\epsilon\frac{m_2 - m_1}{k}\right) + 2(1 - (a_1 + a_2)) \geq 0. \quad (22)$$

It is non-negative, therefore $\mathcal{L}\left(a_1, a_2, \frac{m_1}{k}, \frac{m_2}{k}\right)$ is an increasing function of a_2 reaching its minimum at $a_2 = \pi\frac{m_2^2}{k^2}$. For a_1, we have a concave quadratic function that should reach minimum at one of the boundaries, which are $a_1 = \pi\frac{m_1^2}{k^2}$ and $a_1 = 1 - a_2 = 1 - \pi\frac{m_2^2}{k^2}$. Unfortunately, the minimum could occur at either point depending on the configuration. Therefore, we should consider two cases, which we formulate as the following lemmas, denoting $\frac{m_1}{k} = x$ and $\frac{m_2}{k} = y$.

Lemma 3. *For $0 \leq x, y \leq \frac{1}{2}$ and $\pi x^2 + \pi y^2 \leq 1$ for the parameter $0 < t_\epsilon \leq \frac{4\pi}{1+8\pi} \approx 0.4808$ we have the following inequality:*

$$\mathcal{L}\left(1 - \pi y^2, \pi y^2, x, y\right) \geq 0.$$

Lemma 4. *For $0 \leq x, y \leq \frac{1}{2}$ and $\pi x^2 + \pi y^2 \leq 1$ for the parameter $0 < t_\epsilon \leq \frac{1}{2}$ we have the following inequality:*

$$\mathcal{L}\left(\pi x^2, \pi y^2, x, y\right) \geq 0.$$

Combining these we conclude that for $t_\epsilon \leq \frac{4\pi}{1+8\pi}$ the statement of Theorem 1 holds. But this condition is satisfied for all $\epsilon \leq \log(\frac{1+12\pi}{1+4\pi}) \approx 1.04822$, which concludes the proof.
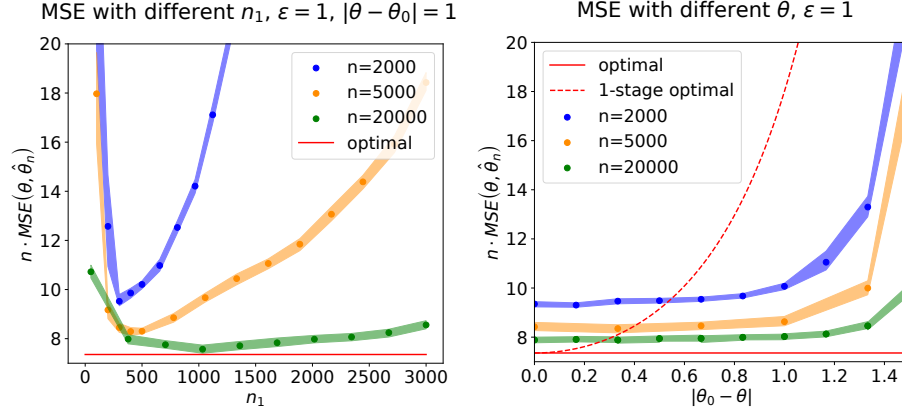
Figure 1: Left panel: Scaled MSE of the private estimator as a function of the number n_1 of first-stage samples. Right panel: Scaled MSE of the private estimator as a function of the initial guess θ_0.

2.2 One-Stage Variance

In this subsection, we address the limitations of using a single-stage procedure. The asymptotic variance of an estimator based on only one stage as in (8) with $n_1 = n$ can be expressed by the following formula:

$$\frac{1}{4}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \frac{1}{\phi(\theta-\theta_0)^2}\left(1-\left(\frac{e^\epsilon-1}{e^\epsilon+1}\right)^2 (1-2\Phi(\theta_0-\theta))^2\right).$$

This expression resembles the optimal asymptotic variance derived in Theorem 1 when $\theta_0 \approx \theta$. A crucial observation is that for single-stage estimation, the asymptotic variance deteriorates exponentially as $|\theta_0 - \theta|$ increases. This motivates the introduction of a two-stage procedure.

Proof. The estimator from the first stage, as described in (8), is updated according to the following formula. We assume that all n samples are used for this estimate, that is, we calculate the asymptotic variance of

$$\tilde{\theta}_n = \theta_0 - \Phi^{-1}\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1}\bar{Z}_n\right). \quad (23)$$

Here, $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$, where $Z_1, \ldots, Z_n$ are iid random variables taking values in $\{-1, 1\}$ according to (5) and $X_i \sim N(\theta, 1)$. To calculate the asymptotic variance of $\tilde{\theta}_n = f(\bar{Z}_n)$ with $f(x) := \theta_0 - \Phi^{-1}\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1}x\right)$, we apply the delta method,

$$\sqrt{n}(f(\bar{Z}_n)-f(\mathbb{E}(Z_1))) \xrightarrow[n\to\infty]{D} N(0, [f'(\mathbb{E}(Z_1))]^2\mathrm{Var}(Z_1)).$$

Thus, we need the expectation and variance of Z_1, which are given by

$$\mathbb{E}(Z_1) = \frac{e^\epsilon-1}{e^\epsilon+1}(1-2\Phi(\theta_0-\theta)),$$

$$\mathrm{Var}(Z_1) = 1-\left(\frac{e^\epsilon-1}{e^\epsilon+1}\right)^2(1-2\Phi(\theta_0-\theta))^2.$$

Since the inverse function theorem gives $f'(x) = \frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1}\left(\phi\circ\Phi^{-1}\left(\frac{1}{2}-\frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1}x\right)\right)^{-1}$ and we have $f(\mathbb{E}[Z_1]) = \theta$, the asymptotic variance of $\tilde{\theta}_n$ is given by

$$\frac{1}{4}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \frac{1}{\phi(\theta_0-\theta)^2}\left(1-\left(\frac{e^\epsilon-1}{e^\epsilon+1}\right)^2(1-2\Phi(\theta_0-\theta))^2\right),$$

which concludes the proof. $\square$

2.3 General variance

In this subsection, we describe how our efficient method for mean estimation can be adjusted to the case of a general, but known, variance. Consider original sensitive data $X_1, \ldots, X_n$ that are independently and identically distributed as $P_{\theta,\sigma^2} = N(\theta, \sigma^2)$ with a known variance $\sigma^2 > 0$. Hence, in this subsection, the statistical model is given by $\mathcal{P}_{\sigma^2} = \{N(\theta, \sigma^2) : \theta \in \mathbb{R}\}$. The only conceptual challenge is to determine a lower bound on the asymptotic variance of any locally private method. Let

$$I_\theta(Q\mathcal{P}_{\sigma^2}) := \mathbb{E}_{Z\sim QP_{\theta,\sigma^2}}\left[\left(\frac{\partial \log QP_{\theta,\sigma^2}(Z)}{\partial\theta}\right)^2\right]$$

denote the information about θ contained in the Q-sanitized data when σ^2 is fixed. In particular, with our previous notation we have $\mathcal{P}_1 = \mathcal{P}$. The following result extends Theorem 1 to the case of general $\sigma^2 > 0$.

Theorem 3. *For $\theta \in \mathbb{R}$, $\sigma^2 > 0$, $\epsilon > 0$ and $Q \in \mathcal{Q}_\epsilon$ we have*

$$I_\theta(Q\mathcal{P}_{\sigma^2}) = \frac{1}{\sigma^2}I_{\frac{\theta}{\sigma}}(Q_\sigma\mathcal{P}),$$

for some mechanism $Q_\sigma \in \mathcal{Q}_\epsilon$ that depends only on σ. In particular, by Theorem 1, if $\epsilon \le 1.04$, it holds that

$$I_\theta(Q\mathcal{P}_\sigma) \le \frac{1}{\sigma^2}I_{\frac{\theta}{\sigma}}\left(Q_{\frac{\theta}{\sigma},\epsilon}^{sgn}\mathcal{P}\right) = \frac{1}{\sigma^2}\frac{2}{\pi}\frac{(e^\epsilon-1)^2}{(e^\epsilon+1)^2}, \quad (24)$$

for all $Q \in \mathcal{Q}_\epsilon$.

In accordance with standard statistical intuition, Theorem 3 now shows that for general known $\sigma^2 > 0$ we should rescale the original data $X_i/\sigma \sim N(\frac{\theta}{\sigma}, 1)$, estimate the ratio $\frac{\theta}{\sigma}$ using the two-stage procedure in (8) and (9) based on the sign-mechanism and then scale back by σ to achieve the optimal variance given by the inverse of the upper bound in (24). It is then clear that this final rescaling by the true σ amounts exactly to the additional factor σ^2 in the asymptotic variance (cf. (24)).

If the variance σ^2 is actually unknown and has to be estimated from appropriately sanitized data, things become much more involved. Estimating the vector (θ, σ^2) efficiently appears to be an ill-posed problem under local differential privacy in the sense that there is no maximizing element Q in the partial Loewner ordering of semidefinite matrices which maximizes the 2×2 Fisher-Information matrix $I_{\theta, \sigma^2}(Q\mathcal{P}^{(2)})$, $\mathcal{P}^{(2)} := \{N(\theta, \sigma^2) : \theta \in \mathbb{R}, \sigma^2 > 0\}$. Intuitively, this can be understood by noticing that in order to estimate both θ and σ^2 we would have to extract two different features of each sensitive X_i in a locally private way and we have to make a decision on how to distribute the privacy budget ϵ over these two tasks, while for estimation of a single parameter, we can use the full privacy budget on that one. Even if σ^2 is only considered to be a nuisance parameter and the goal is still an efficient estimation of θ, it is currently not clear what the appropriate lower bound for the asymptotic variance and an efficient procedure would look like. We defer these important and fundamental issues to future research.

We close this discussion by mentioning that for general unknown $\sigma^2 > 0$ our first stage estimator (8) is not even consistent for θ. However, consistent locally private, yet inefficient, estimators for both θ and σ^2 can be constructed using the method of moments and clipping as explained, for instance, in Theorem 4.3 of Steinberger (2024).

3 Experiments

In this section, we provide numerical experiments to investigate the finite sample performance of our two-stage locally private estimation procedure described in (8) and (9) in dependence on the tuning parameters $\theta_0 \in \mathbb{R}$ and $n_1 \in \mathbb{N}$, for different sample sizes n and $\epsilon = 1$. For the plots, we generated 50000 Montecarlo samples and computed the 95% confidence intervals via bootstrapping. In the left panel of Figure 1, we plot the scaled mean squared error (MSE) of $\hat{\theta}_n$ as a function of the number n_1 of samples in the first stage of our private estimation procedure. From the plots, we see that a proper choice of n_1 is crucial to get a small MSE. Especially too small values of n_1 must be avoided.

In the right panel of Figure 1, we plot the scaled MSE as a function of the difference $\theta_0 - \theta$, where n_1 was chosen
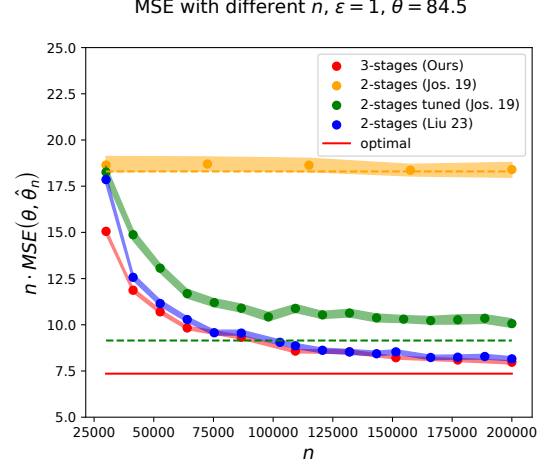

Figure 2: The plot compares the performance of our enhanced 3-stage mechanism with the original mechanism from Joseph et al. (2019), including a tuned version in which we adjusted the number of samples allocated to the first stage. Additionally, it is compared with Liu et al. (2023), which has been enhanced by a preliminary stage from Joseph et al. (2019). The horizontal lines represent the theoretical variance limits for each mechanism.

such that the corresponding MSE function in the left panel is minimized. The results are not surprising. The MSE increases as the initial guess θ_0 moves away from the true value θ. Notice that our results do not provide uniform convergence in θ, and in fact it is known that with differential privacy it is impossible to achieve uniformity over an unbounded parameter space.

3.1 3-Stage Mechanism

Estimating the true mean θ for large initial differences $|\theta - \theta_0|$ is challenging for our two-stage mechanism. To enhance performance, we incorporate the first stage of the mechanism from Joseph et al. (2019), which estimates θ bitwise with high probability, provided an upper bound on the number of bits to estimate. In their approach, the same sign mechanism is used for the second stage. However, our theory predicts that splitting the data equally between these stages is suboptimal. Even when tuned with a better choice of steps allocated to the first stage, their mechanism remains suboptimal because the first stage fails to consistently estimate θ.

We provide additional results of a 3-stage mechanism in Figure 2, with $\theta = 84.5$, under the assumption that the unknown θ is in the range $[0, 128]$. With varying sample sizes n, we observe that the MSE converges to the optimal value for our 3-stage procedure. Specifically, we allocate $n_0 = 15,000$ samples to the preliminary stage based on Joseph et al. (2019), which we find to be close to optimal

for estimating the first 7 bits of θ. We also provide the theoretical limits for the 2-stage mechanisms based on the formula (2.2) We optimized over the choice of n_1 for the 3-stage mechanism and found that for values of n ranging from $30,000$ to $200,000$, the optimal n_1 initially grows from 400 to 700 before almost saturating around 700. We conjecture that as n continues to grow, n_1 might increase logarithmically; however, we lack the theoretical understanding to confirm this. We also compare our approach with the method of Liu et al. (2023), enhanced by the preliminary stage of Joseph et al. (2019). Our results show that for large n, their method performs almost identically to ours, while for smaller n, it may perform slightly worse. The method of Liu et al. (2023) has been proven to achieve a variance that we have shown to be asymptotically optimal. We illustrate its performance in Figure 2 as well.

Our results show that the 3-stage mechanism allows us to estimate θ asymptotically optimally across a broader range of θ for a finite n.

4   Discussion

This study addressed the problem of estimating the unknown mean θ of a unit variance Gaussian distribution under local differential privacy (LDP). We proved that the sign mechanism, which computes the sign of the difference $X_i - \theta$ and applies randomized response, has the lowest asymptotic variance among all locally differentially private mechanisms in the high-privacy regime ($\epsilon \leq 1.04$). However, this mechanism is purely theoretical since it requires knowledge of θ, which we are trying to estimate. To address this issue, we propose a two-stage mechanism that allows for the estimation of θ without any assumptions. For practical considerations, we enhanced our mechanism with an additional pre-step from Joseph et al. (2019) to provide a better initial guess for θ, enabling a good private estimate over a wide range of θ with a finite sample size n.

However, extending our theoretical approach to more complex scenarios presents challenges. For $\epsilon \geq 2$, the binary mechanism becomes provably suboptimal, as evidenced by the Fisher Information, which converges to $\frac{2}{\pi}$ as $\epsilon \to \infty$, rather than to 1, which is the Fisher Information in classical (non-private) estimation. Identifying exact bounds or closed form expressions of optimal mechanisms for larger ϵ remains an open question. Additionally, extending our method to the multidimensional case is difficult, as optimizing the inverse Fisher Information matrix is ill-posed. The current theoretical framework also does not apply to minimizing the sum of variances in this context. Future research will need to develop new approaches to address these challenges and fully characterize optimal LDP mechanisms in more complex settings.

References

Asi, H., Vitaly, F., Jelani, N., Huy, N., and Kunal, T. (2023). Fast optimal locally private mean estimation via random projections. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Asi, H., Vitaly, F., and Kunal, T. (2022). Optimal algorithms for mean estimation under local differential privacy. In *International Conference on Machine Learning (ICML)*.

Barnes, L. P., Chen, W.-N., and Özgür, A. (2020). Fisher information under local differential privacy. *IEEE Journal on Selected Areas in Information Theory*.

Berry, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society (AMS)*.

Dinur, I. and Nissim, K. (2003). Revealing information while preserving privacy. In *ACM Symposium on Principles of Database Systems (PODS)*.

Duchi, J. and Rogers, R. (2019). Lower bounds for locally private estimation via communication complexity. In *Conference on Learning Theory (COLT)*.

Dwork, C. (2008). Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation (TAMC)*.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*.

Dwork, C. and Nissim, K. (2004). Privacy-preserving datamining on vertically partitioned databases. In *Annual International Cryptology Conference (CRYPTO)*.

Evfimievski, A., Gehrke, J., and Srikant, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *ACM Symposium on Principles of Database Systems (PODS)*.

Joseph, M., Kulkarni, J., Mao, J., and Wu, S. Z. (2019). Locally private Gaussian estimation. *Advances in Neural Information Processing Systems (NeurIPS)*.

Kairouz, P., Oh, S., and Viswanath, P. (2016). Extremal mechanisms for local differential privacy. *The Journal of Machine Learning Research (JMLR)*.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer Science & Business Media.

Liu, Y., Hu, Q., Ding, L., and Kong, L. (2023). Online local differential private quantile inference via self-normalization. In *International Conference on Machine Learning (ICML)*.

Nam, S.-H. and Lee, S.-H. (2022). A tighter converse for the locally differentially private discrete distribution estimation under the one-bit communication constraint. *IEEE Signal Processing Letters*.

Steinberger, L. (2024). Efficiency in local differential privacy. *The Annals of Statistics*, 52(1):2139–2166. arXiv:2301.10600.

van der Vaart, A. W. (2007). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, S., Huang, L., Wang, P., Nie, Y., Xu, H., Yang, W., Li, X.-Y., and Qiao, C. (2016). Mutual information optimally local private discrete distribution estimation. *arXiv preprint arXiv:1607.08025*.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*.

Ye, M. and Barg, A. (2018). Optimal schemes for discrete distribution estimation under locally differential privacy. *IEEE Transactions on Information Theory*.

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] The two-stage procedure works in linear time. The required sample size is thoroughly discussed in the experimental section.

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [No] We believe the implementation is straightforward, and all experiments can be easily recreated using the detailed descriptions provided in the manuscript.

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). [Yes]

   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes] We provide a full description of the experimental details, including data splits and hyperparameters, necessary for reproducibility.

   (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). [Yes]

   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [No] We compute our experiments on CPU, those experiments do not require any specific hardware.

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. [Yes]

   (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 Appendix

A.1 Proof of Theorem 2

Rather than relying on Theorem 4.12 in Steinberger (2024) and checking all their assumptions, we here present a direct proof that also has the advantage of being self-contained. We begin by showing the consistency of first and second-stage estimators. For convenience, fix $h \in \mathbb{R}$ and set $\theta_n := \theta + h/\sqrt{n}$. All probabilities, expectations, variances and stochastic convergence results below are with respect to R_{θ_n}.

A.1.1 Consistency

Showing consistency of the first-stage estimator (8) is rather straight forward, because $Z_1, \ldots, Z_{n_1}$ are iid with values in $\{-1, 1\}$ and

$$\mathbb{E}[Z_1] = P(Z_1 = 1) - P(Z_1 = -1) = 2P(Z_1 = 1) - 1$$
$$= 2\frac{e^\epsilon}{1 + e^\epsilon} - 2\frac{e^\epsilon - 1}{e^\epsilon + 1}\Phi(\theta_0 - \theta_n) - 1 = \frac{e^\epsilon - 1}{e^\epsilon + 1}(1 - 2\Phi(\theta_0 - \theta_n)).$$

Thus, by Markov's inequality, $\bar{Z}_{n_1}$ converges in probability to $\frac{e^\epsilon - 1}{e^\epsilon + 1}(1 - 2\Phi(\theta_0 - \theta)) \in (-\frac{e^\epsilon - 1}{e^\epsilon + 1}, \frac{e^\epsilon - 1}{e^\epsilon + 1})$, and thus $P(|\bar{Z}_{n_1}| < \frac{e^\epsilon - 1}{e^\epsilon + 1}) \to 1$ as $n \to \infty$ and $\tilde{\theta}_{n_1}$ converges to θ in probability, by the continuous mapping theorem.

Consistency of (9) can be shown in a similar way, using the conditional Markov inequality. In view of consistency of $\tilde{\theta}_{n_1}$ it suffices to show that $\bar{Z}_{n_2} = \frac{1}{n_2}\sum_{i=n_1+1}^{n} Z_i \to 0$ in probability as $n \to \infty$. Conditionally on $\tilde{\theta}_{n_1}$, the $Z_{n_1+1}, \ldots, Z_n$ are iid. The conditional expectation is computed in the same way as above, that is

$$\mathbb{E}[\bar{Z}_{n_2}|\tilde{\theta}_{n_1}] = \mathbb{E}[Z_n|\tilde{\theta}_{n_1}] = \frac{e^\epsilon - 1}{e^\epsilon + 1}(1 - 2\Phi(\tilde{\theta}_{n_1} - \theta_n)),$$

and it converges to zero in probability as $n \to \infty$. The Z_i's only take values 1 or -1 and thus their conditional variance is bounded by 1. Now

$$P(|\bar{Z}_{n_2}| > \varepsilon) \leq P(|\bar{Z}_{n_2} - \mathbb{E}[\bar{Z}_{n_2}|\tilde{\theta}_{n_1}]| + |\mathbb{E}[\bar{Z}_{n_2}|\tilde{\theta}_{n_1}]| > \varepsilon)$$
$$\leq \mathbb{E}\left[P\left(|\bar{Z}_{n_2} - \mathbb{E}[\bar{Z}_{n_2}|\tilde{\theta}_{n_1}]| > \frac{\varepsilon}{2}\Big|\tilde{\theta}_{n_1}\right)\right] + P\left(|\mathbb{E}[\bar{Z}_{n_2}|\tilde{\theta}_{n_1}]| > \frac{\varepsilon}{2}\right)$$
$$\leq \frac{4}{\varepsilon^2}\mathbb{E}\left[1 \wedge \text{Var}[\bar{Z}_{n_2}|\tilde{\theta}_{n_1}]\right] + P\left(|\mathbb{E}[Z_n|\tilde{\theta}_{n_1}]| > \frac{\varepsilon}{2}\right)$$
$$\leq \frac{4}{\varepsilon^2 n_2} + P\left(|\mathbb{E}[Z_n|\tilde{\theta}_{n_1}]| > \frac{\varepsilon}{2}\right) \xrightarrow{n \to \infty} 0.$$

A.1.2 Asymptotic normality

For asymptotic normality, first notice that we already showed above that $|\bar{Z}_{n_1}|$ and $|\bar{Z}_{n_2}|$ are bounded by $\frac{e^\epsilon - 1}{e^\epsilon + 1}$ with asymptotic probability one. Hence, it suffices to proceed on the event where both of these absolute values are appropriately bounded. Write $G(p) := \Phi^{-1}(p)$ and consider the scaled estimation error of (9), that is

$$\sqrt{n}(\hat{\theta}_n - \theta_n) = -\sqrt{n}\left(\Phi^{-1}\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon + 1}{e^\epsilon - 1}\bar{Z}_{n_2}\right) - (\tilde{\theta}_{n_1} - \theta_n)\right)$$
$$= -\sqrt{n}\left(G\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon + 1}{e^\epsilon - 1}\bar{Z}_{n_2}\right) - G(\Phi(\tilde{\theta}_{n_1} - \theta_n))\right).$$

By a first order Taylor expansion and writing $A_n := \frac{1}{2} - \frac{1}{2}\frac{e^\epsilon + 1}{e^\epsilon - 1}\bar{Z}_{n_2} = \frac{1}{n_2}\sum_{i=n_1+1}^{n}\left(\frac{1}{2} - \frac{1}{2}\frac{e^\epsilon + 1}{e^\epsilon - 1}Z_i\right)$ and $B_n := \Phi(\tilde{\theta}_{n_1} - \theta_n)$, we get

$$\sqrt{n}(G(A_n) - G(B_n)) = G'(B_n)\sqrt{n}(A_n - B_n) + \frac{1}{2}\sqrt{n}\int_{B_n}^{A_n} G''(t)(A_n - t)^2 dt. \tag{25}$$

From our results of the previous subsection we see that $A_n \to \frac{1}{2}$ and $B_n \to \frac{1}{2}$ in probability, and since G' is continuous on $[0, 1]$ we have $G'(B_n) \to G'(\frac{1}{2}) = [\phi \circ \Phi^{-1}(\frac{1}{2})]^{-1} = \sqrt{2\pi}$, in probability. Next, we show that $\sqrt{n}(A_n - B_n) \to$

$N(0, \frac{1}{4}(\frac{e^\epsilon+1}{e^\epsilon-1})^2)$ in distribution. For $i = n_1 + 1, \ldots, n$, define $Y_i := \frac{1}{2} - \frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1}Z_i - \Phi(\tilde{\theta}_{n_1} - \theta_n)$ and note that conditional on $\tilde{\theta}_{n_1}$ they are iid with conditional mean zero and conditional variance

$$\text{Var}[Y_i|\tilde{\theta}_{n_1}] = \frac{1}{4}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2\left[1 - \left(\frac{e^\epsilon-1}{e^\epsilon+1}(1 - 2\Phi(\tilde{\theta}_{n_1}-\theta_n))\right)^2\right] \xrightarrow[n\to\infty]{i.p.} \frac{1}{4}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2.$$

Moreover, $\text{Var}[Y_i|\tilde{\theta}_{n_1}] \geq \frac{1}{4}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2[1 - (\frac{e^\epsilon-1}{e^\epsilon+1})^2]$ and $\mathbb{E}[|Y_i|^3|\tilde{\theta}_{n_1}] \leq \frac{3}{2} + \frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1}$. Thus, from the Berry-Esseen bound (cf. Berry, 1941, Theorem 1) applied to the conditional distribution, we get for every $t \in \mathbb{R}$,

$$\left| P\left(\sqrt{n_2}\frac{A_n - B_n}{\sqrt{\text{Var}(Y_n|\tilde{\theta}_{n_1})}} \leq t\right) - \Phi(t)\right| = \left| P\left(\sqrt{n_2}\sum_{i=n_1+1}^{n}\frac{Y_i}{\sqrt{\text{Var}(Y_i|\tilde{\theta}_{n_1})}} \leq t\right) - \Phi(t)\right|$$

$$\leq \mathbb{E}\left| P\left(\sqrt{n_2}\sum_{i=n_1+1}^{n}\frac{Y_i}{\sqrt{\text{Var}(Y_i|\tilde{\theta}_{n_1})}} \leq t\Big|\tilde{\theta}_{n_1}\right) - \Phi(t)\right| \leq \frac{1.88(\frac{3}{2} + \frac{1}{2}\frac{e^\epsilon+1}{e^\epsilon-1})}{\frac{1}{8}\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^3[1 - (\frac{e^\epsilon-1}{e^\epsilon+1})^2]^{\frac{3}{2}}\sqrt{n_2}}.$$

Applying Slutski's theorem and noting that $n/n_2 \to 1$, we get the desired convergence of $\sqrt{n}(A_n - B_n) \to N(0, \frac{1}{4}(\frac{e^\epsilon+1}{e^\epsilon-1})^2)$ in distribution. Hence, $G'(B_n)\sqrt{n}(A_n - B_n) \to N(0, \frac{\pi}{2}(\frac{e^\epsilon+1}{e^\epsilon-1})^2)$. From Theorem 1 we see that this is the optimal asymptotic variance. Thus, the proof is finished if we can show that the remainder term in (25) converges to zero in probability. But this is easily seen, because on the event where $A_n, B_n \in [\frac{1}{4}, \frac{3}{4}]$, which has asymptotic probability one, the remainder term is bounded in absolute value by

$$\frac{1}{2}|\sqrt{n}(A_n - B_n)||A_n - B_n|^2 \sup_{t\in[\frac{1}{4},\frac{3}{4}]}|G''(t)| \xrightarrow[n\to\infty]{i.p.} 0.$$

A.2 Proof of Lemma 3

We need to show that the inequality in Lemma 3 holds, i.e., that

$$L(x, y, t_\epsilon) := \mathcal{L}\left(1 - \pi y^2, \pi y^2, x, y\right) = 2t_\epsilon(2\pi y^2 - 1) + 4t_\epsilon^2(y - x)(2\pi y^2 - 1) - 4t_\epsilon^2(y - x)^2 + 1 \geq 0. \qquad (26)$$

This is a quadratic function in x and we can express it as

$$L(x, y, t_\epsilon) = -4t_\epsilon^2 x^2 - 4xt_\epsilon^2(2\pi y^2 - 1 - 2y) + 2t_\epsilon(2\pi y^2 - 1) + 4t_\epsilon^2(2\pi y^2 - 1)y - 4t_\epsilon^2 y^2 + 1.$$

The minimal value is achieved on the tails, for $x = 0$ or $x = \frac{1}{2}$. First, we consider the case when $x = 0$.

A.2.1 Case $x = 0$

$$L(0, y, t_\epsilon) = 2t_\epsilon(2\pi y^2 - 1) + 4t_\epsilon^2(2\pi y^2 - 1)y - 4t_\epsilon^2 y^2 + 1. \qquad (27)$$

We will prove that for $t_\epsilon \leq \frac{4\pi}{1+8\pi} < \frac{1}{2}$ the expression in (27) is greater than 0. This expression is a polynomial of the third power in terms of y, and we minimize it on the segment $0 \leq y \leq \frac{1}{2}$. For $y = 0$ we have

$$L(0, 0, t_\epsilon) = 1 - 2t_\epsilon \geq 0, \qquad (28)$$

which is true for $t_\epsilon \leq \frac{1}{2}$. For $y = \frac{1}{2}$, we have:

$$L(0, 1/2, t_\epsilon) = \pi t_\epsilon - 2t_\epsilon + 2t_\epsilon^2\left(\frac{\pi}{2} - 1\right) + 1 - t_\epsilon^2 \geq 0, \qquad (29)$$

which is true for $t_\epsilon \leq 1$. Next, we consider the potential optimum in between where the derivative is equal to 0:

$$\frac{\partial L(0, y, t_\epsilon)}{\partial y} = 8t_\epsilon \pi y + 24t_\epsilon^2 \pi y^2 - 4t_\epsilon^2 - 8t_\epsilon^2 y \tag{30}$$

$$= 4t_\epsilon \left(2\pi y + 6t_\epsilon \pi y^2 - t_\epsilon - 2t_\epsilon y\right) \tag{31}$$

$$= 4t_\epsilon \left(6t_\epsilon \pi y^2 + y(2\pi - 2t_\epsilon) - t_\epsilon\right) = 0. \tag{32}$$

By solving the quadratic equation we obtain an optimal value

$$y^*(t_\epsilon) = \frac{t_\epsilon - \pi + \sqrt{6\pi t_\epsilon^2 + (\pi - t_\epsilon)^2}}{6\pi t_\epsilon}. \tag{33}$$

The other root is negative whereas the root in (33) is between 0 and $\frac{1}{2}$ since

$$\pi - t_\epsilon \le \sqrt{6\pi t_\epsilon^2 + (\pi - t_\epsilon)^2} \le 3\pi t_\epsilon + (\pi - t_\epsilon). \tag{34}$$

Next, we will prove that the function $L(0, y^*(t_\epsilon), t_\epsilon)$ is a decreasing function of t_ϵ. To that end, consider first the derivative of y^*,

$$\frac{\partial y^*(t_\epsilon)}{\partial t_\epsilon} = \frac{1}{6t_\epsilon^2} \left[1 - \frac{(\pi t_\epsilon^{-1} - 1)}{\sqrt{6\pi + (\pi t_\epsilon^{-1} - 1)^2}}\right] > 0, \tag{35}$$

and

$$\frac{\partial L(0, y^*, t_\epsilon)}{\partial t_\epsilon} = 2(2\pi(y^*)^2 - 1) + 8t_\epsilon \pi y^* \frac{\partial y^*}{\partial t_\epsilon} + 8t_\epsilon(2\pi(y^*)^2 - 1)y^* + 24t_\epsilon^2 \pi(y^*)^2 \frac{\partial y^*}{\partial t_\epsilon}$$

$$+ 4t_\epsilon^2(2\pi(y^*)^2 - 1)\frac{\partial y^*}{\partial t_\epsilon} - 8t_\epsilon(y^*)^2 - 8t_\epsilon^2 y^* \frac{\partial y^*}{\partial t_\epsilon}$$

$$= (2\pi(y^*)^2 - 1)(2 + 8t_\epsilon y) - 8t_\epsilon(y^*)^2 + \frac{\partial y^*}{\partial t_\epsilon} \left[8\pi t_\epsilon y^* + 24\pi t_\epsilon^2(y^*)^2 - 4t_\epsilon^2 - 8t_\epsilon^2 y^*\right]$$

$$= (2\pi(y^*)^2 - 1)(2 + 8t_\epsilon y^*) - 8t_\epsilon(y^*)^2 + 4t_\epsilon \left[2\pi y^* + 6\pi t_\epsilon(y^*)^2 - t_\epsilon - 2t_\epsilon y^*\right] \frac{\partial y^*}{\partial t_\epsilon}$$

Now we can substitute $y^*(t_\epsilon)$, namely $2\pi(y^*)^2 = \frac{1}{3} - \frac{2y^*}{3t_\epsilon}(\pi - t_\epsilon)$ and $2\pi y^* + 6\pi t_\epsilon(y^*)^2 - t_\epsilon - 2t_\epsilon y^* = 0$ to obtain

$$\frac{\partial L(0, y^*(t_\epsilon), t_\epsilon)}{\partial t_\epsilon} = \left(-\frac{2}{3} - \frac{2y^*}{3t_\epsilon}(\pi - t_\epsilon)\right)(2 + 8t_\epsilon y^*) - 8t_\epsilon(y^*)^2 < 0 \tag{36}$$

Thus the function $t_\epsilon \mapsto L(0, y^*(t_\epsilon), t_\epsilon)$ is continuous and strictly decreasing. Therefore it suffices to identify the point where it reaches zero. We will argue that it is when $t_\epsilon = \frac{4\pi}{1+8\pi}$. For $t_\epsilon = \frac{4\pi}{1+8\pi}$, we have

$$y^* = \frac{\frac{4\pi}{1+8\pi} - \pi + \sqrt{6\pi \left(\frac{4\pi}{1+8\pi}\right)^2 + \left(\pi - \frac{4\pi}{1+8\pi}\right)^2}}{6\pi \frac{4\pi}{1+8\pi}} = \frac{3\pi - 8\pi^2 + \sqrt{96\pi^3 + (8\pi^2 - 3\pi)^2}}{24\pi^2} = \frac{1}{4\pi}. \tag{37}$$

By substituting it into the $L(0, y^*(t_\epsilon), t_\epsilon)$ we get

$$L(0, y^*(t_\epsilon), t_\epsilon)\Big|_{t_\epsilon = \frac{4\pi}{1+8\pi}} = \left[2t_\epsilon(2\pi(y^*)^2 - 1) + 4t_\epsilon^2(2\pi(y^*)^2 - 1)y^* - 4t_\epsilon^2(y^*)^2 + 1\right]_{t_\epsilon = \frac{4\pi}{1+8\pi}} \tag{38}$$

$$= \frac{8\pi}{1+8\pi}\left(\frac{1}{8\pi} - 1\right) + \frac{16\pi}{(1+8\pi)^2}\left(\frac{1}{8\pi} - 1\right) - \frac{4}{(1+8\pi)^2} + 1 \tag{39}$$

$$= \frac{1 - 64\pi^2}{(1+8\pi)^2} + \frac{2 - 16\pi}{(1+8\pi)^2} - \frac{4}{(1+8\pi)^2} + 1 \tag{40}$$

$$= \frac{1 - 64\pi^2 + 2 - 16\pi - 4 + 1 + 16\pi + 64\pi^2}{(1+8\pi)^2} = 0. \tag{41}$$

Therefore, the original inequality holds when $t_\epsilon \le \frac{4\pi}{1+8\pi}$, which concludes the proof of the case $x = 0$.

A.2.2   Case $x = \frac{1}{2}, 0 \le y \le \sqrt{\frac{1}{\pi} - \frac{1}{4}}$

We need to show the following inequality for the function $L(x, y, t_\epsilon)$ defined in (26):

$$L(1/2, y, t_\epsilon) = 2t_\epsilon(2\pi y^2 - 1) + 4t_\epsilon^2(y - 1/2)(2\pi y^2 - 1) - 4t_\epsilon^2(y - 1/2)^2 + 1 \ge 0. \tag{42}$$

First, we note that for $y = 0$,

$$L(1/2, 0, t_\epsilon) = -2t_\epsilon + 2t_\epsilon^2 + 1 - t_\epsilon^2 = (t_\epsilon - 1)^2 \ge 0. \tag{43}$$

Next, for $y = \sqrt{\frac{1}{\pi} - \frac{1}{4}}$, we have

$$L\left(\frac{1}{2}, \sqrt{\frac{1}{\pi} - \frac{1}{4}}, t_\epsilon\right) = -t_\epsilon(\pi - 2) + 2t_\epsilon^2 \left(\frac{1}{2} - \sqrt{\frac{1}{\pi} - \frac{1}{4}}\right)(\pi - 2) + 1 - 4t_\epsilon^2\left(\sqrt{\frac{1}{\pi} - \frac{1}{4}} - \frac{1}{2}\right)^2. \tag{44}$$

To simplify the expression, let us denote $\alpha = \sqrt{\frac{1}{\pi} - \frac{1}{4}}$. Then the desired inequality is reduced to

$$L\left(\frac{1}{2}, \alpha, t_\epsilon\right) = 2t_\epsilon^2\left(\frac{1}{2} - \alpha\right)(\pi - 3 + 2\alpha) - t_\epsilon(\pi - 2) + 1 \ge 0. \tag{45}$$

The minimum of this quadratic function is achieved when $t_\epsilon = \frac{\pi - 2}{2(1 - 2\alpha)(\pi - 3 + 2\alpha)} \approx 1.8$, therefore it is a decreasing function of t_ϵ on the range $t_\epsilon \le 1$ and $L(1/2, \alpha, 1) \approx 0.175 > 0$, therefore the inequality (45) is fulfilled for $t_\epsilon \le 1$. Now we need to compute the derivative of the function with respect to y to find a possible minimal value on the segment $0 \le y \le \sqrt{\frac{1}{\pi} - \frac{1}{4}}$.

$$\frac{\partial L(1/2, y, t_\epsilon)}{\partial y} = 8t_\epsilon \pi y + 4t_\epsilon^2(2\pi y^2 - 1) + 16\pi t_\epsilon^2(y - 1/2)y - 8t_\epsilon^2(y - 1/2) \tag{46}$$

$$= 4t_\epsilon\left(2\pi y + t_\epsilon 2\pi y^2 - t_\epsilon + 4\pi t_\epsilon y^2 - 2\pi t_\epsilon y - 2t_\epsilon y + t_\epsilon\right) \tag{47}$$

$$= 8t_\epsilon\left(3\pi t_\epsilon y^2 + (\pi - \pi t_\epsilon - t_\epsilon)y\right) = 0 \tag{48}$$

The case when $y = 0$ we have already considered, the second root is

$$y = \frac{t_\epsilon(\pi + 1) - \pi}{3\pi t_\epsilon}. \tag{49}$$

For $0 < t_\epsilon \le \frac{1}{2}$ this fraction is always negative. Therefore, the minimum is achieved when $y = 0$ or $y = \sqrt{\frac{1}{\pi} - \frac{1}{4}}$, concluding the proof of the lemma.

A.3   Proof of Lemma 4

We need to show that

$$L(x, y, t_\epsilon) := \mathcal{L}\left(\pi x^2, \pi y^2, x, y\right) = 2t_\epsilon \pi(y^2 - x^2) + 4t_\epsilon^2 \pi(y - x)(y^2 - x^2) - \pi^2(x^2 + y^2)^2$$
$$+ 2\pi(x^2 + y^2) - 4t_\epsilon^2(y - x)^2 \ge 0.$$

We consider the function L of two variables (x, y) and parameter t_ϵ. We minimize this function to prove that the minimal value is achieved when $x = y = 0$ and therefore the inequality holds. We start by proving that for $t_\epsilon \le \frac{1}{2}$ the optimal value has to be on the boundary, that is, we show that the optimum can not be achieved in the interior by showing that the partial derivatives cannot simultaneously be equal to 0. Consider the system of derivatives

$$\begin{cases} \frac{\partial L(x, y, t_\epsilon)}{\partial y} = 4t_\epsilon \pi y + 8t_\epsilon^2 \pi(y^2 - x^2) + 4t_\epsilon^2 \pi(y - x)^2 - 4\pi^2(x^2 + y^2)y + 4\pi y - 8t_\epsilon^2(y - x) = 0, \\ \frac{\partial L(x, y, t_\epsilon)}{\partial x} = -4t_\epsilon \pi x - 8t_\epsilon^2 \pi(y^2 - x^2) + 4t_\epsilon^2 \pi(y - x)^2 - 4\pi^2(x^2 + y^2)x + 4\pi x + 8t_\epsilon^2(y - x) = 0. \end{cases}$$

We transform this system into the equivalent system of the sum and difference of those equalities, namely,

$$\begin{cases} 4t_\epsilon\pi(y-x) + 8t_\epsilon^2\pi(y-x)^2 - 4\pi^2(x^2+y^2)(y+x) + 4\pi(x+y) = 0, \\ 4t_\epsilon\pi(y+x) + 16t_\epsilon^2\pi(y^2-x^2) - 4\pi^2(x^2+y^2)(y-x) + 4\pi(y-x) - 16t_\epsilon^2(y-x) = 0. \end{cases} \tag{50}$$

For simplicity, we divide everything by 4π and collect terms to get

$$\begin{cases} t_\epsilon(y-x)(1+2t_\epsilon(y-x)) + (y+x)(1-\pi(x^2+y^2)) = 0, \\ t_\epsilon(y+x)(1+4t_\epsilon(y-x)) + (y-x)(1-\pi(x^2+y^2)) - \frac{4}{\pi}t_\epsilon^2(y-x) = 0. \end{cases} \tag{51}$$

From the first equation, we conclude that $y - x < 0$, or otherwise the sum cannot be zero as $(y+x)(1-\pi(x^2+y^2)) > 0$ in the interior of the domain. From the first equation, we can express $1 - \pi(x^2+y^2)$ as

$$1 - \pi(x^2+y^2) = \frac{t_\epsilon(x-y)(1+2t_\epsilon(y-x))}{x+y}, \tag{52}$$

substitute it in the second equation and multiply by $x + y > 0$ to get

$$t_\epsilon(y+x)^2(1+4t_\epsilon(y-x)) - t_\epsilon(x-y)^2(1+2t_\epsilon(y-x)) + \frac{4}{\pi}t_\epsilon^2(x^2-y^2) = 0. \tag{53}$$

Now we divide everything by $t_\epsilon > 0$ and recombine the terms to get

$$G(x,y,t_\epsilon) := t_\epsilon(x-y)\left(2(x-y)^2 - 4(x+y)^2 + \frac{4}{\pi}(x+y)\right) + (x+y)^2 - (x-y)^2 = 0. \tag{54}$$

We will show that for $t_\epsilon \leq \frac{1}{2}$ this identity can not hold. In particular, we show that the left side is larger than 0. To that end, let us consider the change of variables $a = x - y > 0$, $b = x + y \geq a$ and

$$G(x(a,b),y(a,b),t_\epsilon) = t_\epsilon a\left(2a^2 - 4b^2 + \frac{4}{\pi}b\right) + b^2 - a^2 \tag{55}$$

Consider the derivative with respect to b,

$$\frac{\partial G(x(a,b),y(a,b),t_\epsilon)}{\partial b} = 2b + t_\epsilon a\left(\frac{4}{\pi} - 8b\right) = 2b(1 - 4t_\epsilon a) + \frac{4}{\pi}t_\epsilon a > 0 \tag{56}$$

for $t_\epsilon \leq \frac{1}{2}$ and because $a \leq \frac{1}{2}$ by assumption. Therefore $G(x(a,b),y(a,b),t_\epsilon)$ is minimized when b is minimized, but $b \geq a$, thus

$$G(x(a,b),y(a,b),t_\epsilon) \geq t_\epsilon a\left(2a^2 - 4a^2 + \frac{4}{\pi}a\right) = t_\epsilon a^2\left(\frac{4}{\pi} - 2a\right) > 0. \tag{57}$$

Therefore the identity (54) can not hold and there is no optimum in the interior of the domain. Thus, we need to check the boundaries, which leads us to five cases:

1. $x^2 + y^2 = \frac{1}{\pi}$:

$$L\left(x, \sqrt{\frac{1}{\pi} - x^2}, t_\epsilon\right) = 2t_\epsilon\pi(y^2 - x^2)(1 + 2t_\epsilon(y-x)) - 4t_\epsilon^2(y-x)^2 + 1$$

$$\geq -\frac{3\pi}{2}|y^2 - x^2| - (y-x)^2 + 1$$

$$\geq -\frac{3\pi}{2}\left(\frac{1}{2} - \frac{1}{\pi}\right) - \left(\frac{1}{2} - \sqrt{\frac{1}{\pi} - \frac{1}{4}}\right)^2 + 1 \geq 0.$$

2. $x = \frac{1}{2}, 0 \leq y \leq \sqrt{\frac{1}{\pi} - \frac{1}{4}}$:

$$L(1/2, y, t_\epsilon) = 4t_\epsilon^2\left(y - \frac{1}{2}\right)^2\left(\pi\left(\frac{1}{2} + y\right) - 1\right) + 2t_\epsilon\pi\left(y^2 - \frac{1}{4}\right) - \pi^2\left(\frac{1}{4} + y^2\right)^2 + 2\pi\left(\frac{1}{4} + y^2\right)$$

$$\geq -\frac{\pi}{4} + \pi\left(\frac{1}{4} + y^2\right) \geq 0.$$

3. $y = \frac{1}{2}, 0 \leq x \leq \sqrt{\frac{1}{\pi} - \frac{1}{4}}$:

$$
L(x, 1/2, t_\epsilon) = 4t_\epsilon^2 \left(x - \frac{1}{2}\right)^2 \left(\pi \left(\frac{1}{2} + x\right) - 1\right) + 2t_\epsilon \pi \left(\frac{1}{4} - x^2\right)
$$
$$
- \pi^2 \left(\frac{1}{4} + x^2\right)^2 + 2\pi \left(\frac{1}{4} + x^2\right) \geq 0.
$$

4. $x = 0$:

$$
\begin{aligned}
L(0, y, t_\epsilon) &= 2t_\epsilon \pi y^2 + 4t_\epsilon^2 \pi y^3 - \pi^2 y^4 + 2\pi y^2 - 4t_\epsilon^2 y^2 \\
&= y^2 (-\pi^2 y^2 + 4t_\epsilon^2 \pi y + 2\pi t_\epsilon + 2\pi - 4t_\epsilon^2) \\
&\geq y^2 \left(-\frac{\pi^2}{4} + 2\pi - 1\right) \geq 0.
\end{aligned}
$$

5. $y = 0$:

$$
\begin{aligned}
L(x, 0, t_\epsilon) &= -2t_\epsilon \pi x^2 + 4t_\epsilon^2 \pi x^3 - \pi^2 x^4 + 2\pi x^2 - 4t_\epsilon^2 x^2 \\
&= x^2 (-\pi^2 x^2 + 4t_\epsilon^2 \pi x - 2t_\epsilon \pi + 2\pi - 4t_\epsilon^2) \\
&\geq x^2 (-\frac{\pi^2}{4} + 2\pi t_\epsilon^2 - 2t_\epsilon \pi + 2\pi - 4t_\epsilon^2) \geq x^2 \left(-\frac{\pi^2}{4} + \frac{\pi}{2} - \pi + 2\pi - 1\right) \geq 0.
\end{aligned}
$$

This concludes the proof.

A.4 Proof of Theorem 3

Let $(\mathcal{Z}, \mathcal{G})$ denote the measurable space on which the mechanism Q generates its outputs, let p_{θ,σ^2} be the density of the normal distribution with mean θ and variance σ^2 and write $s_{\theta|\sigma^2}(x) = \frac{x-\theta}{\sigma^2}$ for the score in the model $\mathcal{P}_{\sigma^2}$. In view of Lemma 3.1 and Lemma D.2, the Fisher-Information can be represented by

$$
I_\theta(Q\mathcal{P}_{\sigma^2}) = \int_{\mathcal{Z}} t_{\theta|\sigma^2}(z)^2 q_{\theta|\sigma^2}(z) \nu(dz), \tag{58}
$$

where $\nu(dz) = Q(dz|x_0)$ for some $x_0 \in \mathbb{R}$, $q_{\theta|\sigma^2}(z) := \int_{\mathbb{R}} q(z|x) p_{\theta,\sigma^2}(x) dx$ is a ν-density of the sanitized data $Z_i \sim QP_{\theta,\sigma^2}$, $t_{\theta|\sigma^2}(z) := \int_{\mathbb{R}} s_{\theta|\sigma^2}(x) \frac{q(z|x)p_{\theta,\sigma^2}(x)}{q_{\theta|\sigma^2}(z)} dx$ is the score at θ in the corresponding statistical model QP_{σ^2} and $q : \mathcal{Z} \times \mathbb{R} \to [e^{-\epsilon}, e^\epsilon]$ is a measurable function such that $q(z|x)p_{\theta,\sigma^2}(x)$ is a $\nu \otimes \lambda$-density of $Q(dz|x)P_{\theta,\sigma^2}(dx)$. By a simple change of variable, we get

$$
q_{\theta|\sigma^2}(z) = \int_{\mathbb{R}} q(z|\sigma x) p_{\frac{\theta}{\sigma},1}(x) dx
$$

and

$$
t_{\theta|\sigma^2}(z) = \frac{1}{\sigma} \int_{\mathbb{R}} \left(x - \frac{\theta}{\sigma}\right) \frac{q(z|\sigma x) p_{\frac{\theta}{\sigma},1}(x)}{q_{\theta|\sigma^2}(z)} dx = \frac{1}{\sigma} t_{\frac{\theta}{\sigma},1}(z).
$$

Hence, if we define $Q_\sigma(A|x) := Q(A|\sigma x)$, we see that $Q_\sigma \in \mathcal{Q}_\epsilon$ and that $q(z|\sigma x)p_{\frac{\theta}{\sigma},1}(x)$ is a $\nu \otimes \lambda$-density of $Q_\sigma(dz|x)P_{\frac{\theta}{\sigma},1}(dx)$. In particular, $QP_{\theta,\sigma^2} = Q_\sigma P_{\frac{\theta}{\sigma},1}$ and $t_{\frac{\theta}{\sigma},1}$ is a score at $\frac{\theta}{\sigma}$ in the model $Q_\sigma \mathcal{P}$. Therefore, we recognize (58) to be equal to $\frac{1}{\sigma^2} I_{\frac{\theta}{\sigma}}(Q_\sigma \mathcal{P})$. $\qquad \square$