

---

# Distribution-Aware Mean Estimation under User-level Local Differential Privacy

---

**Corentin Pla**  
CREST-ENSAE  
Criteo AI Lab  
Paris, France

**Maxime Vono**  
Criteo AI Lab  
Paris, France

**Hugo Richard**  
Criteo AI Lab  
Paris, France

## Abstract

We consider the problem of mean estimation under user-level local differential privacy, where  $n$  users are contributing through their local pool of data samples. Previous work assume that the number of data samples is the same across users. In contrast, we consider a more general and realistic scenario where each user  $u \in [n]$  owns  $m_u$  data samples drawn from some generative distribution  $\mu$ ;  $m_u$  being unknown to the statistician but drawn from a known distribution  $M$  over  $\mathbb{N}^*$ . Based on a distribution-aware mean estimation algorithm, we establish an  $M$ -dependent upper bounds on the worst-case risk over  $\mu$  for the task of mean estimation. We then derive a lower bound. The two bounds are asymptotically matching up to logarithmic factors and reduce to known bounds when  $m_u = m$  for any user  $u$ .

## 1 Introduction

In the current artificial intelligence era, machine learning (ML) techniques and algorithms have become common tools for data analysis and decision making. In most cases, their empirical success is based on the availability of large training datasets, which is for instance theoretically justified by data probabilistic modelling and the Bernstein-von Mises approximation asserting that under appropriate conditions, the more data can be processed, the more accurate the inference can be performed (Bardenet et al., 2017; Bottou et al., 2018; Cornish et al., 2019). However, in the meantime,

growing user concerns and regulatory frameworks have emerged regarding how user personal data could be collected and processed by third-party entities. These data privacy issues hence have incentivised users of machine learning applications to ask for methods preserving data anonymity (Shokri and Shmatikov, 2015; Abadi et al., 2016; Zaeem and Barber, 2020).

Whilst many privacy-preserving tools have been proposed (Garcelon et al., 2022; Knott et al., 2021; Ohri-menko et al., 2016; McMahan et al., 2017a; Vono et al., 2022), the *workhorse* framework in machine learning is differential privacy (DP). DP allows to formally protect user’s privacy in a quantifiable manner while maintaining the statistical utility. It dates back at least to Warner (1965) but has received a renewed interest after the work of Dwork et al. (2014). As an example, DP has been implemented in real-world scenario by large technology companies such as Meta, Apple, or LinkedIn (Erlingsson et al., 2014a; Tang et al., 2017; Messing et al., 2020; Rogers et al., 2021); public bodies such as the United States Census Bureau (Abowd, 2018); or non-governmental organisations such as Wikimedia Foundation (Adeleye et al., 2023). The DP framework, which requires outputs of a *differentially-private* algorithm to be indistinguishable when a single contribution changes, has been formalised first in the so-called central model where data is aggregated before being privatised (Dwork et al., 2006; Hardt and Talwar, 2010). The main ingredient to ensure such guarantees is to randomise seminal algorithms by incorporating calibrated noise. To embrace more stringent privacy guarantees, DP has then evolved into a distributed setting, called *local* differential privacy (LDP), which privatises data before aggregation, alleviating the need to resort to a trusted aggregator (Kasiviswanathan et al., 2010).

In contrast to the central setting, LDP comes with a degraded privacy/utility trade-off mainly due to noise involved in each local data contribution. Such fundamental limits have been established by a series of

recent works, focusing on specific tasks and associated algorithms. For instance, the problem of finding optimal private protocols has already been addressed for mean estimation (Duchi et al., 2018), density estimation (Butucea et al., 2019), and hypothesis testing (Berrett and Butucea, 2020).

Traditionally, differentially-private models based on LDP assume that users are contributing with only one data sample (Duchi et al., 2018). Whilst this framework is sufficient for basic use-cases such as mean wage estimation of several individuals, it cannot be applied to a broader set of real-world applications where users may contribute via multiple data points. As seminal examples, users can contribute with a local database made of multiple item (*e.g.*, product or movie) ratings in order to train recommendation systems (McSherry and Mironov, 2009); or can hold several sequences of words typed on their mobile device keyboards which are then used to train next-word prediction models (McMahan et al., 2017c). Such general applications nourished a research area focusing on *user-level* differential privacy, where each user holds more than one data sample and seeks to maintain the privacy of her entire collection. User-level DP has been considered under both central and local settings for statistical tasks already pointed out previously, that is mean estimation, empirical risk minimisation, sparse estimation, non-parametric density estimation (Girgis et al., 2022; Acharya et al., 2022b; Ma et al., 2024; Kent et al., 2024; Levy et al., 2021). Unfortunately, the user-level DP paradigm considered in aforementioned state-of-the-art research work lacks realism as it involves an unrealistic assumption of uniformity over the number of samples. More precisely, current work assume that the number of data samples provided by each user is the same, which does not hold in many real-world settings as outlined, for instance, by federated learning applications (McMahan et al., 2017b; Wang et al., 2021).

This paper aims at filling this gap. Among statistical tasks that have been tackled in the literature, *univariate mean estimation* stands for the primary and most interesting problem to consider as it is the basis to many other protocols. Indeed, associated results can be extended to multivariate mean estimation via specific data transformations involving Hadamard matrices (Chen et al., 2020), non-parametric density estimation via the trigonometric basis (Butucea et al., 2019), sparse mean estimation (Acharya et al., 2022b) or regression via stochastic convex optimisation (Bassily and Sun, 2023).

As such, we consider in this paper the problem of univariate mean estimation under user-level local differential privacy where each user  $u \in [n]$  locally holds

a dataset of  $m_u \in \mathbb{N}^*$  independent and identically distributed observations from an unknown probability distribution  $\mu$ ;  $m_u$  being drawn from a known discrete probability distribution  $M$ . To the best of our knowledge, we are the first to investigate this general user-level local DP setting.

**Contribution.** Our contribution is three-folded.

- We propose a novel and more realistic user-level LDP framework that allows users to contribute via local data sets of heterogeneous cardinality.
- We instantiate this framework for the task of univariate mean estimation and propose an associated algorithm, coined *distribution-aware mean estimation* (DAME). The proposed algorithm works in two phases: (i) the *localisation* phase aims at finding a bin where the true mean parameter lies with high probability; (ii) the *estimation* phase projects data samples onto the bin to reduce the range. Compared to other variants proposed in previous work, the novelty of DAME lies in (i) the selection of users and the bin size, which are adapted to the data set size distribution  $M$ ; (ii) the estimation phase where the mean is shrunk towards the center of the bin to include data contributions from all users even those having few samples.
- We derive non-asymptotic guarantees on the worst-case risk over  $\mu$  for mean estimation. More precisely, we derive upper bounds via the proposed algorithm DAME, and also provide lower bounds on the same quantity. Based on the derived bounds, we show that DAME is optimal in many scenarios: (i) asymptotically in  $n$  up to a log-factor; (ii) in the item-level ( $m_u = 1$  for any  $u \in [n]$ ) LDP setting, up to constant factor; (iii) in the homogeneous ( $m_u = m \in \mathbb{N}^*$  for any  $u \in [n]$ ) user-level LDP setting, up to a log factor; and (iv) in many regimes interpolating the item-level and user-level settings. The optimality of the derived bounds is verified numerically for several choices of the discrete distribution  $M$ .

**Related Work.** As outlined previously, private mean estimation under the DP paradigm has been investigated in both the central (Dwork et al., 2006; Hardt and Talwar, 2010) and local settings (Duchi et al., 2018). These seminal research works have focused on providing item-level (in contrast to user-level) privacy under the assumption that the number of samples is the same across all users. When the data samples are discrete, the more general problem of frequency estimation has also been well-studied under the same assumptions in both central (Dwork et al.,

2006) and local DP models (Hsu et al., 2012; Erlingson et al., 2014a; Acharya et al., 2019). For instance, Liu et al. (2020) studied the problem of learning discrete distributions with different number of samples per user. However, the aforementioned works used a sub-optimal approach to tackle this user heterogeneity by discarding samples of some users and using the median number of other users to apply known estimation techniques derived in the homogeneous user setting.

Regarding private mean estimation, at least two recent work considered different number of samples per user but under central DP. More precisely, Cummings et al. (2024) took inspiration from private federated learning settings and proposed a private mean estimation algorithm where users first compute local mean estimates which are then aggregated and privatised by a central aggregator. In contrast to our proposed approach, their methodology is non-interactive and very different in terms of algorithmic steps – leading to difficulties to translate their central DP work to our local DP setting. In addition, Zhao et al. (2024) recently proposed a novel private mean estimation approach, based on Huber loss minimisation, which aims at tackling the potential bias of the two-phase procedure we consider in our algorithm DAME.

## 2 Problem Formulation

**Notation.** We set  $[n] = \{1, \dots, n\}$ ,  $[n_1, n_2] = \{n_1, \dots, n_2\}$  and use the notation  $z_{i:j} = (z_i, \dots, z_j)^\top$  and  $Z^{(i:j)} = (Z^{(i)}, \dots, Z^{(j)})^\top$ . We define  $\mathcal{B}([-1, 1])$  as the Borel set over  $[-1, 1]$  and let  $\mu^{\otimes i}$  denote the  $i$ -th product distribution. Through this paper, we will use the mathematical notation,  $\mathcal{O}(\cdot)$ ,  $\Omega(\cdot)$  and  $\Theta(\cdot)$  to describe the complexities of the computation required by private mean estimation algorithms. We recall that  $f(d) = \mathcal{O}(g(d))$  (resp.  $f(d) = \Omega(g(d))$ ) if there exists  $c > 0$  such that  $f(d) \leq cg(d)$  (resp.  $f(d) \geq cg(d)$ ). We use  $f(d) = \Theta(g(d))$  if there exist  $c_1, c_2 > 0$  such that  $c_1g(d) \leq f(d) \leq c_2g(d)$ . We define  $a \wedge b = \min(a, b)$  and  $a \vee b = \max(a, b)$ . Finally we denote  $D_{TV}(\cdot, \cdot)$  the total variance distance and  $D_{KL}(\cdot, \cdot)$  the Kullback-Leibler divergence.

**Framework.** We study the setting of user-level local differential privacy, in which users do not trust the central data aggregator. We consider an environment involving  $n \in \mathbb{N}^*$  users holding *private* local data sets  $\{X^{(u)} = \{X_1^{(u)}, \dots, X_{m_u}^{(u)}\}\}_{u \in [n]}$  of different sizes  $\{m_u\}_{u \in [n]}$ . For any  $u \in [n]$ ,  $m_u$  is distributed according to a discrete distribution  $M$  defined on  $(\mathbb{N}^*, 2^{\mathbb{N}^*})$ , and  $X^{(u)}$  stands for a random variable distributed according to a probability distribution  $\mu$  defined on a measurable space  $(\mathcal{X}, \mathfrak{X})$ . For the sake of simplicity and similar to previous work,

we assume, for any  $u \in [n]$  and  $t \in [m_u]$ , that  $X_t^{(u)} \in [-1, 1]$ . Note that in contrast to *item-level* LDP where  $m_u = 1$  for any  $u \in [n]$ , the considered *user-level* LDP setting assumes that each user  $u \in [n]$  contributes  $m_u \in \mathbb{N}^*$  samples to the global (artificial) dataset so that  $\mathcal{X} = \bigcup_{m=1}^{\infty} [-1, 1]^m$ . As a result, each user local data set is sampled independently and identically according to the joint distribution  $\nu_\mu$  described by the following generative model:

$$m_u \sim M, X^{(u)} \mid m_u = i \sim \mu^{\otimes i}. \quad (1)$$

In order to preserve their private local datasets, users do not directly reveal  $X^{(u)}$  to the statistician but instead send to the latter an obfuscated view  $Z^{(u)}$  of  $X^{(u)}$ , defined over  $(\mathcal{Z}, \mathfrak{Z})$ . More precisely, the random variables  $\{Z^{(u)}\}_{u \in [n]}$  stand for differentially-private versions of  $\{X^{(u)}\}_{u \in [n]}$ . We assume that users sequentially reveal  $\{Z^{(u)}\}_{u \in [n]}$  to the statistician. This assumption means that for any user  $u \in [n]$ ,  $Z^{(u)}$  and  $\{X^{(u')}\}_{u' \neq u}$  are independent given  $X^{(u)}$  and  $Z^{(1:u-1)}$ . Such independence property implies that the conditional probability distribution  $Q$  of  $Z^{(1:n)}$  given  $X^{(1:n)}$ , referred to as *mechanism*, is fully characterised by  $\{Q_u\}_{u \in [n]}$  where, for any  $u \in [n]$ ,  $Q_u$  is the local conditional probability distribution of  $Z^{(u)}$  given  $X^{(u)}, Z^{(1:u-1)}$ , coined *local channel*.

**Local Differential Privacy.** For a given scalar privacy parameter  $\alpha > 0$  and any  $u \in [n]$ , we say that the random variable  $Z^{(u)}$  is an  $\alpha$ -*differentially locally private* view of  $X^{(u)}$  if, for any  $z_{1:u-1} \in \mathcal{Z}^{u-1}$  and  $x, x' \in \mathcal{X}$ , the conditional probability distribution  $Q_u : \mathfrak{Z} \times (\mathcal{X} \times \mathcal{Z}^{u-1}) \rightarrow [0, 1]$  of  $Z^{(u)}$  given  $X^{(u)}, Z^{(1:u-1)}$  satisfies:

$$\sup_{S \in \mathfrak{Z}} \frac{Q_u(S \mid x, z_{1:u-1})}{Q_u(S \mid x', z_{1:u-1})} \leq \exp(\alpha). \quad (2)$$

Note that the previous definition does not constrain  $Z^{(u)}$  to be a data release exclusively based on  $X^{(u)}$ : the local channel  $Q_u$  may be interactive, that is differing based on previous data releases  $Z^{(1:u-1)}$ . When for any  $u \in [n]$ ,  $Z_u$  is an  $\alpha$ -differentially locally private view of  $X_u$ , we say that  $Q = \{Q_u\}_{u \in [n]}$  is an  $\alpha$ -LDP *mechanism*. We refer to  $\mathcal{Q}_\alpha$  as the set of  $\alpha$ -LDP mechanisms.

**Private Mean Estimation.** The objective of mean estimation under user-level LDP is to estimate  $\theta = \mathbb{E}_{X \sim \mu}[X]$  by choosing (i) an  $\alpha$ -LDP mechanism  $Q$  that will generate data  $Z^{(1)}, \dots, Z^{(n)}$  from  $X^{(1)}, \dots, X^{(n)}$ , and (ii) an estimator  $\hat{\theta} : \mathcal{Z}^n \rightarrow [0, 1]$ . The difficulty of the such estimation task is measured by a metric referred to as the *worst-case risk over  $\mu$* , and defined

by:

$$R_{\alpha,n,M} = \inf_{Q \in \mathcal{Q}_\alpha, \hat{\theta} \in \Theta} \sup_{\mu \in \mathcal{D}} \mathbb{E} \left[ \left| \hat{\theta} \left( Z^{(1:n)} \right) - \theta \right|^2 \right]. \quad (3)$$

This paper presents in Section 3 a practical algorithm, referred to as DAME, to perform private mean estimation. In Section 4, we derive upper and lower bounds on the worst-case risk (3) – the upper bound being reached by the proposed algorithm.

### 3 Distribution-Aware Mean Estimation (DAME)

In this section, we present the proposed algorithm to perform private mean estimation under user-level LDP. We refer to the latter as distribution-aware mean estimation (DAME) algorithm, to emphasize that it works under user data set size heterogeneity.

**High-level Description.** Similar to other interactive algorithms proposed in previous work (Berrett and Butucea, 2020; Acharya et al., 2022b,a; Kent et al., 2024; Berrett and Butucea, 2020), DAME proceeds in a two-phase procedure. In the first phase, coined the *localisation phase*, private data of the first half of users is discretised and privatised via a calibrated randomised response mechanism (Warner, 1965). This phase aims to identify a candidate bin where the true mean lies with high probability. In the second phase, coined the *estimation phase*, private data of remaining users is projected onto the candidate bin (slightly enlarged) and privatised via the Laplace mechanism (Dwork et al., 2014, Definition 3.3) where the added noise scales with the width of the enlarged candidate bin. The novelty of DAME is that it adapts the bin size to the data set size distribution  $M$ , and uses a novel biasing/debiasing procedure in the estimation phase to include data from all users even those that have few samples. The following paragraphs provide additional details regarding the two phases involved in DAME.

**Algorithmic Details.** More precisely, we assume an even number of users  $n$ , discarding one user if we need to, and use the first  $n/2$  users for the localisation phase and the last  $n/2$  for the estimation phase. The interval  $[-1, 1]$  is partitioned into non-overlapping sub-intervals  $I_j$  (also called *bins*) of width  $2\tau$ , where  $\tau > 0$  is given by:

$$\tau = \sqrt{\frac{2 \log(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}{\tilde{m}}}, \quad (4)$$

with  $\tilde{m} \in \mathbb{N}^*$  being the *effective maximum data set size* and standing as an hyper-parameter of the DAME. We denote by  $l$  the bin index containing the true mean  $\theta$ , that is  $\theta \in I_l$ .

① In the localisation phase, each user  $u \in [n/2]$  such that her number of samples  $m_u$  verifies  $m_u \geq \tilde{m}$  computes her sample mean  $\bar{X}_{m_u}^{(u)} = (1/m_u) \sum_{t=1}^{m_u} X_t^{(u)}$ , identifies in which sub-intervals  $\bar{X}_{m_u}^{(u)}$  falls into, and computes the indicator vector  $V^{(u)} = (V_j^{(u)})_{j \in [\lceil \frac{1}{\tau} \rceil]}$  defined, for any  $j \in [\lceil \frac{1}{\tau} \rceil]$ , by

$$V_j^{(u)} = \mathbb{1} \left\{ \bar{X}_{m_u}^{(u)} \in \cup_{k \in \{j-1, j, j+1\}} I_k \right\}. \quad (5)$$

Informally, we say that user  $u$  *votes* for the  $j$ -th bin if  $V_j^{(u)} = 1$ . A user therefore votes for the bins that either contain its empirical mean or are neighbors of the bin containing its empirical mean. This ensures that  $I_l$ , the interval containing the true mean  $\theta$ , receives a vote with high probability even when  $\theta$  is close or equal to the border of the bin. The indicator vector of users  $u \in [n/2]$  with  $m_u \leq \tilde{m}$  is given by  $V^{(u)} = 0$ . In other words, users with low number of samples are not allowed to take part to the vote.

Then, a privatised version  $\tilde{V}^{(u)}$  of  $V^{(u)}$  is computed using the randomised response mechanism for binary vectors with three non-zero entries Erlingsson et al. (2014b). More precisely, for any  $j \in [\lceil \frac{1}{\tau} \rceil]$ ,  $\tilde{V}_j^{(u)}$  is defined as:

$$\tilde{V}_j^{(u)} = \begin{cases} V_j^{(u)} & \text{with proba. } \pi_{\alpha/6} = \frac{e^{\alpha/6}}{1+e^{\alpha/6}}, \\ 1 - V_j^{(u)} & \text{otherwise.} \end{cases} \quad (6)$$

The candidate bin  $I_{\hat{j}}$  is the one that receives the highest number of votes, that is:

$$\hat{j} = \operatorname{argmax}_{j \in [\lceil \frac{1}{\tau} \rceil]} \sum_{u=1}^{n/2} \tilde{V}_j^{(u)}. \quad (7)$$

② In the estimation phase, each remaining user  $u \in [n/2+1, n]$  computes an estimate of her respective local empirical mean that is shrunk towards the mid-point  $s_{\hat{j}}$  of  $I_{\hat{j}}$ , and defined as:

$$\hat{X}_{\hat{j}}^{(u)} = \frac{\sqrt{m_u \wedge \tilde{m}}}{\sqrt{\tilde{m}}} \left( \bar{X}_{m_u}^{(u)} + \left( \frac{\sqrt{\tilde{m}}}{\sqrt{m_u \wedge \tilde{m}}} - 1 \right) s_{\hat{j}} \right).$$

Then,  $\hat{X}_{\hat{j}}^{(u)}$  is projected onto  $[L_{\hat{j}}, U_{\hat{j}}]$ , an enlarged copy of  $I_{\hat{j}}$ , where  $6\tau$  are added to the left and right borders to ensure that shrunk estimates are not affected by the projection, with high probability. A Laplace noise scaling with the width of  $[L_{\hat{j}}, U_{\hat{j}}]$  is then added to preserve privacy. For any  $u \in [n/2+1, n]$ , the resulting local mean estimate for user  $u$  writes:

$$\hat{\theta}^{(u)} = \Pi_{\hat{j}} \left( \hat{X}_{\hat{j}}^{(u)} \right) + \frac{14\tau}{\alpha} \ell_u, \quad (8)$$

where  $\Pi_{\hat{j}}(\cdot)$  is the projection onto  $[L_{\hat{j}}, U_{\hat{j}}]$  and  $\ell_u$  is a Laplace random variable.



By debiasing, the estimate of the true mean  $\theta$  is finally given by:

$$\hat{\theta} = \frac{\bar{\theta}\sqrt{\tilde{m}} - \sum_{i=1}^{\tilde{m}} (\sqrt{\tilde{m}} - \sqrt{i}) M(i) s_{\hat{j}}}{\mathbb{E}_{m \sim M} (\sqrt{m \wedge \tilde{m}})}, \quad (9)$$

where  $\bar{\theta} = \frac{2}{n} \sum_{u=\frac{n}{2}+1}^n \theta^{(u)}$  and  $M(i) = \mathbb{P}_{m \sim M}(m = i)$ .

**Algorithm.** The main steps of the proposed algorithm DAME are depicted in Algorithm 1.

## 4 Theoretical Analysis

In this section, we present an upper bound, obtained via DAME detailed in Algorithm 1, on the worst-case risk  $R_{\alpha,n,M}$  defined in (3). Furthermore, we also derive a lower bound on  $R_{\alpha,n,M}$ . These two non-asymptotic bounds allow us to discuss the asymptotic optimality of our results, while drawing connections with known results from the private mean estimation literature.

### 4.1 Lower and Upper Bounds on $R_{\alpha,n,M}$

**Assumption.** In order to derive our non-asymptotic theoretical results, we consider the following assumptions on the main parameters characterising the considered distribution-aware user-level LDP paradigm.

**H1** (Finite expectation  $M$ ). *The data set size distribution  $M$  is known and such that  $\mathbb{E}_{m \sim M}[m] < \infty$ .*

**H2** (Bounded support of  $\mu$ ). *The common data distribution  $\mu$  admits a known support defined as  $[-1, 1]$ .*

**H3** (High privacy regime for  $\alpha$ ). *The LDP parameter  $\alpha > 0$  is set such that  $\alpha \leq 22/35$ .*

**H1** is a rather weak assumption satisfied by most discrete distributions encountered in practice. **H2** is a common assumption in local differential privacy made for instance in Duchi et al. (2018); Bassily (2019); Blasiok et al. (2019). In the worst-case, the square dependency on the range cannot be avoided but for some distributions, better results can be obtained (Bun and Steinke, 2019). Lastly, the high privacy regime in **H3** is a rather strong assumption that we make for the sake of simplicity. This assumption mainly allows to re-use results derived in Duchi et al. (2018), that also rely on **H3**. We believe our results could be extended to lower privacy regimes but leave this task to future work.

**Lower bound.** Theorem 1 below provides a lower bound on the worst-case risk  $R_{\alpha,n,M}$  defined in (3).

**Theorem 1** (Lower bound). *Assume **H1-3**. Then, there exist  $c_1, c_2 > 0$ , independent of  $\alpha, n$  and  $m$ , such*

---

**Algorithm 1** Distribution-Aware Mean Estimation (DAME). The algorithm is described from a centralised point of view. However, the statistician never has access to user data, unless explicitly shared.

---

**Input:** number of users ( $n$ ), privacy parameter ( $\alpha$ ), effective maximum data set size ( $\tilde{m}$ ) and distribution of data set size ( $M$ ).

Compute  $\tau$  according to (4).

Divide  $[-1, 1]$  into non-overlapping intervals  $\{I_j = [l_j, u_j]\}_{j \in [\lceil \frac{1}{\tau} \rceil]}$  of size  $2\tau$ .

▷ **Localisation phase:** users with large data sets elect privately as candidate the bin most likely to contain  $\theta$ .

for  $u \in [n/2]$  do

if  $m_u \geq \tilde{m}$  then

    Compute  $\bar{X}_{m_u}^{(u)} = \frac{1}{m_u} \sum_{t=1}^{m_u} X_t^{(u)}$  .

    Compute  $V_j^{(u)} = \mathbb{1}\left\{\bar{X}_{m_u}^{(u)} \in \cup_{k \in \{j-1, j, j+1\}} I_k\right\}$   
    for any  $j \in [\lceil \frac{1}{\tau} \rceil]$  .

else

    Set  $V^{(u)} = 0$

    Share  $\tilde{V}_j^{(u)} = \begin{cases} V_j^{(u)} & \text{w.p. } \frac{e^{\alpha/6}}{1+e^{\alpha/6}} \\ 1 - V_j^{(u)} & \text{otherwise} \end{cases}; j \in [\lceil \frac{1}{\tau} \rceil]$ .

Compute  $\hat{j} = \operatorname{argmax}_{j \in [\lceil \frac{1}{\tau} \rceil]} \sum_{u=1}^{n/2} \tilde{V}_j^{(u)}$  .

Compute  $L_{\hat{j}} = (l_{\hat{j}} - 6\tau) \vee -1, U_{\hat{j}} = (u_{\hat{j}} + 6\tau) \wedge 1$  and  $s_{\hat{j}} = \frac{l_{\hat{j}} + u_{\hat{j}}}{2}$  .

▷ **Estimation phase:** empirical mean of users with small data sets are biased toward the elected bin.

for  $u \in [\frac{n}{2} + 1, n]$  do

    Set  $\hat{X}_{\hat{j}}^{(u)} = \frac{\sqrt{m_u \wedge \tilde{m}}}{\sqrt{\tilde{m}}} \left( \bar{X}_{m_u}^{(u)} + \left( \frac{\sqrt{\tilde{m}}}{\sqrt{m_u \wedge \tilde{m}}} - 1 \right) s_{\hat{j}} \right)$  .

    Sample  $\ell_u$  from standard Laplace distribution.

    Share  $\hat{\theta}^{(u)} = \Pi_{\hat{j}} \left( \hat{X}_{\hat{j}}^{(u)} \right) + \frac{14\tau}{\alpha} \ell_u$  .

Compute  $\bar{\theta} = \frac{2}{n} \sum_{u=\frac{n}{2}+1}^n \theta^{(u)}$  .

▷ **Bias correction.**

return  $\hat{\theta} = \frac{\bar{\theta}\sqrt{\tilde{m}} - \sum_{i=1}^{\tilde{m}} (\sqrt{\tilde{m}} - \sqrt{i}) M(i) s_{\hat{j}}}{\mathbb{E}_{m \sim M} (\sqrt{m \wedge \tilde{m}})}$  .

---

that the following lower bound holds:

$$R_{\alpha,n,M} \geq \max_{a \in \mathbb{N}} \frac{c_1 e^{-c_2 n \alpha^2 \mathbb{P}_{m \sim M}(m > a)^2}}{n \alpha^2 \mathbb{E}_{m \sim M} [\sqrt{m} \mathbb{1}\{m \leq a\}]^2 \vee 1} . \quad (10)$$

The positive constants  $c_1$  and  $c_2$  are explicitly given in Appendix S1.

Proof sketch, see Appendix S1 for more details.

Consider probability distributions  $\mu_0$  and  $\mu_1$  supported on  $\{-1, 1\}$  such that  $\mu_0(1) = (1 - \delta/2)$  and  $\mu_1(1) = (1 + \delta/2)$ . Consider for  $i \in \{0, 1\}$ ,  $\nu_i = \nu_{\mu_i}$  where  $\nu_{\mu}$  is described in eq. (1). From a sequential application of LeCam's bound, Bretagnolle-Huber (Yu (1997)) and (Duchi et al., 2014, Th 1), it holds that  $R_{n,\alpha,M} \geq \frac{\delta^2}{4} \exp(-12n\alpha^2 D_{TV}(\nu_0, \nu_1)^2)$ . Then Lemma S1 gives  $D_{TV}(\nu_0, \nu_1) \leq \mathbb{E}_{m \sim M}[\sqrt{\frac{m}{2}} D_{KL}(\mu_0, \mu_1) \wedge 1]$  and Lemma S2 gives  $D_{KL}(\mu_0, \mu_1) \leq 3\delta^2$  for  $\delta \in [0, \frac{1}{2}]$ . Then choosing  $\delta^2 = \frac{1}{4(n\alpha^2 \mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}]^2 \vee 1)}$  and optimising over  $a$  gives the lower bound.  $\square$

The lower bound in (10) means that there exists a universal constant  $c_2$  such that given  $M, \alpha, n$ , and for any user-level LDP algorithm returning an estimate  $\hat{\theta}$  of  $\theta = \mathbb{E}_{X \sim \mu}[X]$ , there exists a choice of  $\mu$  such that for any  $a \in \mathbb{N}$ , it holds that

$$\mathbb{E}[|\hat{\theta} - \theta|^2] = \Omega\left(\frac{\exp(-c_2 n \alpha^2 \mathbb{P}_{m \sim M}(m > a)^2)}{n \alpha^2 \mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}]^2 \vee 1}\right).$$

**Upper bound.** Our next result, provided in Theorem 2, analyses the performance of DAME and therefore provides an upper bound on the worst-case risk  $R_{n,\alpha,M}$  defined in (3).

**Theorem 2** (Upper bound). *Assume H1-3. Let  $c_3, c_4, c_5 > 0$  be universal constants, independent of  $n, \alpha$  and  $M$ . Consider the function  $\phi : a \in \mathbb{N}^* \mapsto c_5 / (n \alpha^2 \log[c_4(a n \alpha^2 \vee 1) / \log(c_4(a n \alpha^2 \vee 1))])$  and set  $\tilde{m} = \operatorname{argmax}_{a \in \mathbb{N}^*} \{\mathbb{P}_{m \sim M}(m \geq a)^2 \geq \phi(a) \wedge 1\}$ . Then, we have the following lower bound.*

$$R_{\alpha,n,M} \leq \frac{c_3 \ln(c_4(\sqrt{\tilde{m} n \alpha^2} \vee 1))}{n \alpha^2 (\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}}))^2} \wedge 4. \quad (11)$$

The positive constants  $c_3, c_4$  and  $c_5$  are explicitly given in Appendix S2.

*Proof sketch, see Appendix S2 for more details.* The upper bound is provided by DAME which is instantiated with  $\tilde{m}$  as specified in Theorem 2.  $\tilde{m}$  is defined through  $\phi$  which allows to handle the trade-off between the probability of finding the correct bin in the localization phase and the error term due to the variance. The proof starts by a bias-variance decomposition of the objective and we first focus on the bias. Thanks to Hoeffding bounds (Hoeffding, 1963), each user can estimate  $\theta$  up to a precision scaling in approximately  $1/\sqrt{m_u}$ . Since the bins size scales with  $1/\sqrt{\tilde{m}}$ , users with data set size higher than  $\tilde{m}$  will vote for  $I_l$  or its neighboring bin with high probability. However, because of the privatisation, for each user (including those with low data set size), a bin could gain or lose a vote with probability  $1 - \pi_{\alpha/6}$ .

Lemma S5 in the Appendix precisely upper bounds the probability that the localisation is not successful, i.e.,  $\hat{j}$  is at distance at least 3 from the optimal bin  $l$ . If the localisation is not successful, the error is upper bounded by 4. If the localisation is successful, a user  $u$  with a high number of samples ( $m_u \geq \tilde{m}$ ) is likely to have its empirical mean  $\bar{X}_{m_u}^{(u)}$  in  $[L_{\hat{j}}, U_{\hat{j}}]$  since this interval contains  $\theta$  and is of size  $1/\sqrt{\tilde{m}}$  up to log factors (see Lemma S7). Users with low number of samples are unlikely to be in this interval and therefore their estimate  $\hat{X}_{m_u}^{(u)}$  of the mean is a linear combination of their empirical mean and  $s_{\hat{j}}$  the middle point of bin  $I_{\hat{j}}$  where coefficients are chosen to guarantee that  $\hat{X}_{m_u}^{(u)}$  lies in  $[L_{\hat{j}}, U_{\hat{j}}]$ . This procedure adds a bias that can be computed explicitly and is removed when  $\theta$  is estimated (see eq. (9)) therefore ensuring that the error added by users with low number of samples is controlled (see Lemma S8 and Lemma S10). We then focus on the variance (see Lemma S11) which is given by  $\tilde{m} \operatorname{Var}(\hat{\theta}^{(u)}) / \mathbb{E}_{m \sim M}[\sqrt{m \wedge \tilde{m}}]$  by definition of  $\hat{\theta}$ . The variance of  $\hat{\theta}^{(u)}$  is the sum of the variance of the Laplace noise and the variance of the estimates, the latter being controlled by the projection. The precise formula for  $\tilde{m}$  trades-off the probability of finding the correct bin in the localisation phase and the error terms due to the variance.  $\square$

To reach the upper bound presented in Theorem 2,  $\tilde{m}$  must be set as the solution of the optimisation problem

$$\operatorname{argmax}_{a \in \mathbb{N}^*} \underbrace{\{\mathbb{P}_{m \sim M}(m \geq a)^2 - \phi(a) \wedge 1 \geq 0\}}_{\psi(a)}. \quad (12)$$

Notice that  $\psi$  is decreasing,  $\psi(1) \geq 0$  and  $\psi(\exp(n \alpha^2)) \leq 0$ . It follows that  $\tilde{m}$  can be found via a binary search with less than  $2 \lceil n \alpha^2 \rceil$  iterations.

The optimality of our bounds and comparison with previous work is discussed in Section 4.2.

## 4.2 Discussion

In this section, we demonstrate that our bounds are asymptotically optimal up to log-factors. In addition, we show that they reduce to known bounds in the item-level LDP setting or the homogeneous (i.e.,  $m_u = m \in \mathbb{N}^*$ ) user-level LDP setting. Lastly, we illustrate numerically our theoretical insights by plotting upper and lower bounds for classical distributions and studying the empirical performance of DAME on synthetic data.

**Asymptotic Optimality.** First, we claim that DAME is asymptotically optimal up to a log factor as the number of users  $n$  grows towards infinity. Indeed, by taking  $a \rightarrow \infty$  in Theorem 1, we get the following

Setting	Minimax rate	Reference
$m = 1$	$(n\alpha^2)^{-1}$	Duchi et al. (2018)
$m \in \mathbb{N}^*$	$(mn\alpha^2)^{-1}$	Kent et al. (2024)
$m \sim M$	$(\mathbb{E}_{m \sim M}[\sqrt{m}]^2 n\alpha^2)^{-1}$	DAME (this work)

Table 1: General setting – asymptotic minimax rate with respect to  $\mu$  (as  $n \rightarrow +\infty$ ). Results are given discarding log factors.

lower bound:

$$R_{\alpha,n,M} = \Omega\left(\frac{1}{n\alpha^2 \mathbb{E}_{m \sim M}[\sqrt{m}]^2}\right). \quad (13)$$

Looking at the upper bound in Theorem 2, since, for any  $a \in \mathbb{N}$ ,  $\lim_{n \rightarrow \infty} \phi(a) = 0$ , we get that  $\tilde{m} \rightarrow \infty$ . Furthermore, we have  $\lim_{n \rightarrow \infty} \tilde{m}/(n\alpha^2) \leq 1$ . Indeed,  $\tilde{m} \geq n\alpha^2$  implies that, for any  $x \leq n\alpha^2$ ,  $\mathbb{P}(m \geq x) \geq \sqrt{\phi(x)} \geq \sqrt{\frac{1}{n\alpha^2}}$  leading to  $\mathbb{E}_{m \sim M}[m] \geq \int_0^{n\alpha^2} \mathbb{P}_{m \sim M}(m \geq x) dx \geq \sqrt{n\alpha^2}$ , which would diverge and therefore violate **H1**. Hence, it follows that there exists a universal constant  $c > 0$ , such that:

$$\lim_{n \rightarrow \infty} R_{\alpha,n,M} \left(\frac{\log(n\alpha^2)}{n\alpha^2 \mathbb{E}_{m \sim M}[\sqrt{m}]^2}\right)^{-1} \leq c, \quad (14)$$

where  $R_{\alpha,n,M}$  is defined in (3).

Combining (13) and (14), it follows that the proposed private mean estimation algorithm DAME asymptotically attains tight rates up to log factors. We now show that DAME is tight in settings studied in previous work.

**Comparison with Previous Work - General Setting.** In the item-level LDP setting, the data set size distribution  $M$  is given by  $M(1) = 1$ , where we recall that, for any  $i \in \mathbb{N}^*$ ,  $M(i) = \mathbb{P}_{m \sim M}(m = i)$ . Optimal rates are given by  $\Theta(1/n\alpha^2)$ , see Duchi et al. (2018, Corollary 1). This lower bound coincides with that of Theorem 1 as shown in (13). A naive application of our upper bound in Theorem 2 uses  $\tilde{m} = 1$  and gives a rate of order  $\mathcal{O}(\log(n\alpha^2)/(n\alpha^2))$  which is only tight up to a log factor. However, as argued in the proof of Theorem 2, when  $\tilde{m} = 1$ , there is only one bin equal to the full interval  $[-1, 1]$ . In this case, DAME becomes identical to the optimal procedure derived in Duchi et al. (2018) and therefore reaches the same upper bound.

Table 1 summarises our findings regarding the comparison of our bounds derived in Theorem 1 and Theorem 2 with previous work, in terms of asymptotic optimality.

**Comparison with Previous Work - Toy Example.** We now consider a more involved setting where for  $\rho \in [0, 1]$ ,  $M(m) = \rho, M(1) = (1 - \rho)$ . In this

scenario, taking  $a = 0$  and  $a = m$  in the lower bound defined in Theorem 1 gives:

$$R_{\alpha,n,M} = \Omega\left(\frac{1}{n\alpha^2((1-\rho) + \rho\sqrt{m})^2} \vee \frac{e^{-c_2 n\alpha^2 \rho^2}}{(1-\rho)^2 n\alpha^2 \vee 1}\right). \quad (15)$$

The associated upper bound is given by Corollary 3, detailed below.

**Corollary 3** (Upper bound). *Assume **H1-3**. For  $\rho \in [0, 1]$ , take  $M$  such that  $M(m) = \rho, M(1) = (1 - \rho)$  and assume  $n\alpha^2 \geq 1$ . Then,  $\phi(a) = c_5(n\alpha^2)^{-1} \log[c_4(an\alpha^2)/\log(c_4(an\alpha^2))]$  and we have the following upper bounds:*

$$\begin{aligned} \forall \rho^2 \in [0, 1], R_{n\alpha,M} &= \mathcal{O}\left(\frac{\ln(n\alpha^2)}{n\alpha^2}\right) \\ \forall \rho^2 \geq \phi(m), R_{n,\alpha,M} &= \mathcal{O}\left(\frac{\log(mn\alpha^2)}{n\alpha^2((1-\rho) + \rho\sqrt{m})^2}\right) \\ \forall a \in \{2, \dots, m-1\}, \forall \rho^2 \in [\phi(a), \phi(a+1)], \\ R_{n,\alpha,M} &= \mathcal{O}\left(\frac{\log(an\alpha^2)}{n\alpha^2((1-\rho) + \rho\sqrt{a})^2}\right) \wedge \frac{e^{-\rho^2 \frac{n\alpha^2}{c_5}}}{\rho^2} \end{aligned}$$

*Proof.* For any  $\rho$ , we have  $\tilde{m} \geq 1$  which gives the upper bound in  $\log(n\alpha^2)/(n\alpha^2)$ . When  $\rho^2 \geq \phi(m)$ , then  $\tilde{m} = m$  and the upper bounds follows. Otherwise, the condition on  $\rho^2$  ensures that  $\tilde{m} = a$  which gives the upper bound in  $\log(an\alpha^2)/(\rho^2 an\alpha^2)$ . Then the condition  $\rho^2 \leq \phi(a+1)$  implies  $\rho^2 \leq \phi(2a)$  which leads to  $\log(2c_4 an\alpha^2)/(2c_4 an\alpha^2) \leq \exp(-\rho^2 n\alpha^2/c_5)$ . This implies the bound in  $\exp(-\rho^2 n\alpha^2/c_5)/\rho^2$ .  $\square$

Corollary 3 assumes  $n\alpha^2 \geq 1$ . This is not restrictive as when  $n\alpha^2 \leq 1$  the trivial upper bound 4 matches up to a constant factor the lower bound in  $\Omega(1/(n\alpha^2))$ . When  $\rho \geq \phi(m)$ , the upper and lower bounds derived in (15) and Corollary 3 are tight up to a log factor. In particular, when  $\rho = 1$  and  $\phi(m) \leq 1$ , we get an upper bound in  $\log(mn\alpha^2)/(mn\alpha^2)$  and a matching lower bound up to a log factor. On the other hand, when  $\rho = 1$  and  $\phi(m) \geq 1$ , the upper bound becomes  $\exp(-n\alpha^2/c_5)$  whereas the lower bound scales in  $\exp(c_2 n\alpha^2)$ . Note that  $\rho = 1$  corresponds to the homogeneous user-level setting ( $M(m) = 1$ ). In this setting, the current known upper and lower bounds (Kent et al., 2024, Theorem 6) are respectively given by  $\Omega(\exp(-c'_1 n\alpha^2) \wedge (mn\alpha^2)^{-1})$  and  $\mathcal{O}(\exp(-c'_2 n\alpha^2) + \log(mn\alpha^2)/(mn\alpha^2))$ , where  $c'_1, c'_2$  are universal constants. These known results are again matching our bounds.

Table 2 compares the upper bound of DAME to the upper bounds obtained in Kent et al. (2024).

Regime	DAME
$0 \leq \rho^2 < \phi(2)$	$(n\alpha^2)^{-1}$
$\phi(2) \leq \rho^2 < \phi(3)$	$(\rho\sqrt{2} + (1 - \rho))^2 n\alpha^2)^{-1}$
$\vdots$	$\vdots$
$\phi(a) \leq \rho^2 < 1$	$((\rho\sqrt{a} + (1 - \rho))^2 n\alpha^2)^{-1}$
$1 \leq \rho^2 < \phi(a + 1)$	$(n\alpha^2)^{-1}$

Table 2: Toy example – upper bound on worst-case risk with respect to  $\mu$  when  $M(1) = (1 - \rho)$  and  $M(m) = \rho$ . The guarantees are given discarding log factors. Known bounds in the literature are  $(n\alpha^2)^{-1}$  in all regimes.

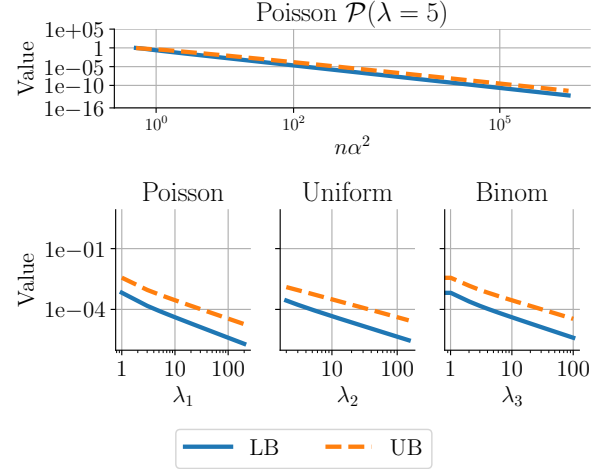
**Empirical study.** We plot in Figure 1a the upper and lower bounds for various distributions which allows us to verify the tightness of our bounds and then benchmark DAME with Duchi et al. (2018) and Kent et al. (2024) on a synthetic dataset.

More precisely, Figure 1a (top) shows the upper bound (dotted orange) and lower bound (blue) when  $M$  is chosen according to either Poisson  $\mathcal{P}(\lambda_1)$ , Uniform  $\mathcal{U}(1, 2\lambda_2 - 1)$  or Binomial  $\mathcal{B}(1000, \lambda_3/1000)$ . We take  $n\alpha^2 = 500$  and vary  $\lambda_1, \lambda_2, \lambda_3$  in a grid of 100 points. We note that the upper bounds and lower bounds form two parallel straight lines in a log-log plot, attesting to the log-tight nature of our bounds. To draw these curves, we need to estimate  $a_1$  resp.  $a_2$  such that  $a_1 = \arg\max_{a \in \mathbb{N}^+} \{\phi(a)\}$  and  $a_2$  maximises the lower bound. The parameter  $a_1$  is found via a binary search and  $a_2$  via a grid search.

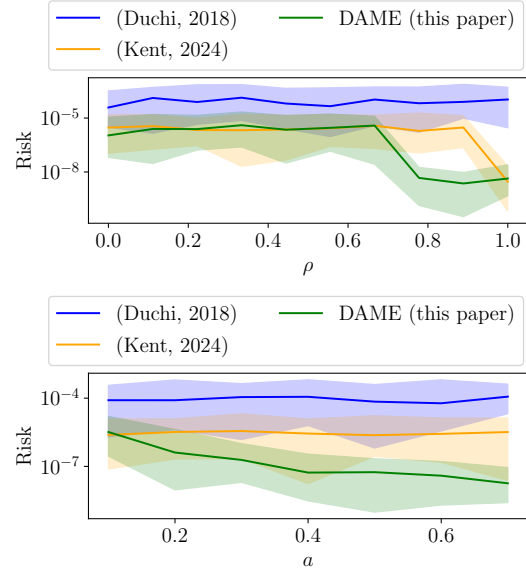
We then fix the distribution  $M$  to a Poisson with parameter  $\lambda = 5$  and vary  $n\alpha^2$  in a grid of 100 points. The results are displayed in Figure 1a (bottom) and show that the upper and lower bounds behave similarly with respect to  $n\alpha^2$ .

Then, we benchmark DAME, Duchi et al. (2018) and Kent et al. (2024) on two synthetic datasets. In both datasets,  $\mu$  is such that  $\mu(1) = \mu(-1) = 0.5$ . In the first dataset, the data set size distribution  $M$  is defined by  $M(m_1) = 1 - \rho$  and  $M(m_2) = \rho$  where  $m_1 = 10^5, m_2 = 10^6$  and  $\rho$  varies on a linear grid with 10 points between 0 and 1. In the second dataset,  $M$  is a power law distribution with density  $x \mapsto \frac{a}{(m_2 - m_1)} \left( \frac{x - m_1}{m_2 - m_1} \right)^{a-1}$  where  $m_1 = 10^5, m_2 = 10^6$  and  $a$  varies on a linear grid with 7 points between 0 and 1. We use  $n = 10^4, \alpha = 22/35$  and  $\tilde{m}$  is chosen via a binary search. The results, displayed in Figure 1b, show that DAME is able to obtain better results than other approaches because it adapts to the distribution, benefiting from the presence of users with many samples even when some users have few of samples.

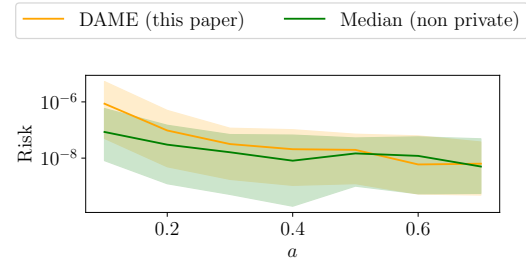
Finally, we compare DAME to a non-private version of



(a) Top: UB & LB for varying  $n\alpha^2$ . Bottom: UB & LB for varying parameters of Poisson, Uniform, and Binomial distributions.



(b) Risk of the estimator given by Duchi et al. (2018), Kent et al. (2024) and DAME for various distributions. Top :  $M(m_1) = 1 - \rho$  and  $M(m_2) = \rho$ . Bottom :  $m \sim \text{Pow}(a)$ .



(c) Risk of the estimator given by DAME and by a non private version (Kent et al., 2024) where only the users with number of samples larger than the median are kept.



Kent et al. (2024), which keeps only users with more samples than the median. Experiments are carried out with  $m$  following a power law and with the same parameters as previously. Unsurprisingly, we note that if each user accurately reveals the number of samples they possess, the estimator performs better. This highlights the privacy cost incurred by keeping the number of samples that each user has private.

The code for reproducing the experiments is available at <https://github.com/hugorichard/Dame>.

## 5 Conclusion

In this paper, we consider the mean estimation problem under user-level LDP constraints in the case where each user  $u \in [n]$  owns  $m_u$  data samples drawn from some generative distribution  $\mu$ ;  $m_u$  being unknown to the statistician but drawn from a known distribution  $M$  over  $\mathbb{N}$ . We propose a mean estimation algorithm, coined **DAME**, and derive non-asymptotic guarantees on the worst-case risk over  $\mu$  presented in Theorem 1 and Theorem 2. These results show that **DAME** provides asymptotic optimality up to some log-factors and matches previously known bounds (Duchi et al., 2018; Kent et al., 2024).

As univariate mean estimation is a sub-routine used in the current best-known algorithms for multivariate mean estimation, sparse estimation, or distribution estimation (Duchi et al., 2014; Acharya et al., 2022a) in item-level LDP, we believe **DAME** can easily be used as a building block to provide efficient algorithms for various tasks in user-level LDP without the need to assume that all users have an identical dataset size – an assumption not met in practice.

An important assumption of this work is the perfect knowledge of  $M$  (**H1**) that is used for the debiasing step of Algorithm 1 and in the choice of  $\tilde{m}$  defined in (12), leading to the upper bound in Theorem 2. Whether inexact knowledge of  $M$  can be used to achieve similar bounds is a question of interest.

Lastly, **DAME** is interactive. Whether there exists a non-interactive algorithm with similar good properties could be the subject of future work.

## References

- Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS ’16, page 308–318, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341394. doi: 10.1145/2976749.2978318.
- John M. Abowd. The u.s. census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD ’18, page 2867, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355520. doi: 10.1145/3219819.3226070. URL <https://doi.org/10.1145/3219819.3226070>.
- Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1120–1129. PMLR, 16–18 Apr 2019. URL <https://proceedings.mlr.press/v89/acharya19a.html>.
- Jayadev Acharya, Clément L Canonne, Ziteng Sun, and Himanshu Tyagi. The role of interactivity in structured estimation. In *Conference on Learning Theory*, pages 1328–1355. PMLR, 2022a.
- Jayadev Acharya, Yuhua Liu, and Ziteng Sun. Discrete distribution estimation under user-level local differential privacy, 2022b. URL <https://arxiv.org/abs/2211.03757>.
- Temilola Adeleye, Skye Berghel, Damien Desfontaines, Michael Hay, Isaac Johnson, Cléo Lemoisson, Ashwin Machanavajjhala, Tom Magerlein, Gabriele Modena, David Pujol, Daniel Simmons-Marengo, and Hal Triedman. Publishing Wikipedia usage data with strong privacy guarantees. *arXiv e-prints*, art. arXiv:2308.16298, August 2023. doi: 10.48550/arXiv.2308.16298.
- Rémi Bardenet, Arnaud Doucet, and Chris Holmes. On Markov chain Monte Carlo methods for tall data. *Journal of Machine Learning Research*, 18(47):1–43, 2017.
- Raef Bassily. Linear queries estimation with local differential privacy. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 721–729. PMLR, 2019.
- Raef Bassily and Ziteng Sun. User-level private stochastic convex optimization with optimal rates. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 1838–1851. PMLR, 23–29 Jul 2023. URL <https://proceedings.mlr.press/v202/bassily23a.html>.

- Thomas B. Berrett and Cristina Butucea. Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms, 2020. URL <https://arxiv.org/abs/2005.12601>.
- Jaroslav Blasiok, Mark Bun, Aleksandar Nikolov, and Thomas Steinke. Towards instance-optimal private query release. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2480–2497. SIAM, 2019.
- Léon Bottou, Frank E. Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018. doi: 10.1137/16M1080173.
- Mark Bun and Thomas Steinke. Average-case averages: Private algorithms for smooth sensitivity and mean estimation. *Advances in Neural Information Processing Systems*, 32, 2019.
- Cristina Butucea, Amandine Dubois, Martin Kroll, and Adrien Saumard. Local differential privacy: Elbow effect in optimal density estimation and adaptation over besov ellipsoids, 2019. URL <https://arxiv.org/abs/1903.01927>.
- Wei-Ning Chen, Peter Kairouz, and Ayfer Ozgur. Breaking the communication-privacy-accuracy trilemma. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 3312–3324. Curran Associates, Inc., 2020. URL [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/222afbe0d68c61de60374b96f1d86715-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/222afbe0d68c61de60374b96f1d86715-Paper.pdf).
- Rob Cornish, Paul Vanetti, Alexandre Bouchard-Cote, George Deligiannidis, and Arnaud Doucet. Scalable Metropolis-Hastings for exact Bayesian inference with large datasets. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1351–1360. PMLR, 09–15 Jun 2019.
- Rachel Cummings, Vitaly Feldman, Audra McMillan, and Kunal Talwar. Mean estimation with user-level privacy under data heterogeneity. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS ’22, Red Hook, NY, USA, 2024. Curran Associates Inc. ISBN 9781713871088.
- John Duchi. *Lecture Notes on Statistics and Information Theory*. Stanford, 2019. URL <https://web.stanford.edu/class/stats311/lecture-notes.pdf>.
- John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy, data processing inequalities, and statistical minimax rates, 2014. URL <https://arxiv.org/abs/1302.3203>.
- John C Duchi, Michael I Jordan, and Martin J Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.
- Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-32732-5.
- Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in TCS*, 9(3–4):211–407, 2014.
- Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS ’14, page 1054–1067, New York, NY, USA, 2014a. Association for Computing Machinery. ISBN 9781450329576. doi: 10.1145/2660267.2660348. URL <https://doi.org/10.1145/2660267.2660348>.
- Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS’14, page 1054–1067. ACM, November 2014b. doi: 10.1145/2660267.2660348. URL <http://dx.doi.org/10.1145/2660267.2660348>.
- Evrard Garcelon, Matteo Pirodda, and Vianney Perchet. Encrypted linear contextual bandit. In Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera, editors, *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 2519–2551. PMLR, 28–30 Mar 2022. URL <https://proceedings.mlr.press/v151/garcelon22a.html>.
- Antonious M. Girgis, Deepesh Data, and Suhas Digavi. Distributed user-level private mean estimation. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2196–2201, 2022. doi: 10.1109/ISIT50566.2022.9834713.
- Moritz Hardt and Kunal Talwar. On the geometry of differential privacy. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC ’10, page 705–714, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781450300506. doi: 10.

- 1145/1806689.1806786. URL <https://doi.org/10.1145/1806689.1806786>.
- Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. ISSN 01621459, 1537274X. URL <http://www.jstor.org/stable/2282952>.
- Justin Hsu, Sanjeev Khanna, and Aaron Roth. Distributed private heavy hitters. In Artur Czumaj, Kurt Mehlhorn, Andrew Pitts, and Roger Wattenhofer, editors, *Automata, Languages, and Programming*, pages 461–472, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-31594-7.
- J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately?, 2010. URL <https://arxiv.org/abs/0803.0924>.
- Alexander Kent, Thomas B. Berrett, and Yi Yu. Rate optimality and phase transition for user-level local differential privacy, 2024. URL <https://arxiv.org/abs/2405.11923>.
- Brian Knott, Shobha Venkataraman, Awni Hanun, Shubho Sengupta, Mark Ibrahim, and Laurens van der Maaten. Crypten: Secure multi-party computation meets machine learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 4961–4973. Curran Associates, Inc., 2021. URL [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/2754518221cfbc8d25c13a06a4cb8421-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/2754518221cfbc8d25c13a06a4cb8421-Paper.pdf).
- Daniel Asher Nathan Levy, Ziteng Sun, Kareem Amin, Satyen Kale, Alex Kulesza, Mehryar Mohri, and Ananda Theertha Suresh. Learning with user-level privacy. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL [https://openreview.net/forum?id=G1jmxFOtY\\_](https://openreview.net/forum?id=G1jmxFOtY_).
- Yuhan Liu, Ananda Theertha Suresh, Felix Xinnan X Yu, Sanjiv Kumar, and Michael Riley. Learning discrete distributions: user vs item-level privacy. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 20965–20976. Curran Associates, Inc., 2020. URL [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/f06edc8ab534b2c7ecbd4c2051d9cb1e-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/f06edc8ab534b2c7ecbd4c2051d9cb1e-Paper.pdf).
- Yuheng Ma, Ke Jia, and Hanfang Yang. Better locally private sparse estimation given multiple samples per user, 2024. URL <https://arxiv.org/abs/2408.04313>.
- Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 20–22 Apr 2017a. URL <https://proceedings.mlr.press/v54/mcmahan17a.html>.
- Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282, Fort Lauderdale, FL, USA, 20–22 Apr 2017b. PMLR.
- H. B. McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. Learning differentially private recurrent language models. In *International Conference on Learning Representations*, 2017c. URL <https://api.semanticscholar.org/CorpusID:3461939>.
- Frank McSherry and Ilya Mironov. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, page 627–636, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605584959. doi: 10.1145/1557019.1557090. URL <https://doi.org/10.1145/1557019.1557090>.
- Solomon Messing, Christina DeGregorio, Bennett Hillenbrand, Gary King, Saurav Mahanti, Zagreb Mukerjee, Chaya Nayak, Nate Persily, Bogdan State, and

- Arjun Wilkins. Facebook Privacy-Protected Full URLs Data Set, 2020. URL <https://doi.org/10.7910/DVN/TDOAPG>.
- Olga Ohrimenko, Felix Schuster, Cédric Fournet, Aastha Mehta, Sebastian Nowozin, Kapil Vaswani, and Manuel Costa. Oblivious multi-party machine learning on trusted processors. In *Proceedings of the 25th USENIX Conference on Security Symposium*, SEC’16, page 619–636, USA, 2016. USENIX Association. ISBN 9781931971324.
- Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kancha, Shraddha Sahay, and Parvez Ahammad. LinkedIn’s audience engagements api: A privacy preserving data analytics system at scale. *Journal of Privacy and Confidentiality*, 11, 12 2021. doi: 10.29012/jpc.782.
- Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- R. Shokri and V. Shmatikov. Privacy-preserving deep learning. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 909–910, 2015. doi: 10.1109/ALLERTON.2015.7447103.
- Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and XiaoFeng Wang. Privacy loss in apple’s implementation of differential privacy on macos 10.12. *CoRR*, abs/1709.02753, 2017. URL <http://dblp.uni-trier.de/db/journals/corr/corr1709.html#abs-1709-02753>.
- Maxime Vono, Vincent Plassier, Alain Durmus, Aymeric Dieuleveut, and Eric Moulines. QLSD: Quantised Langevin stochastic dynamics for Bayesian federated learning. In *AISTATS*, 2022.
- Jianyu Wang, Zachary Charles, Zheng Xu, Gauri Joshi, H. Brendan McMahan, Blaise Aguera y Arcas, Maruan Al-Shedivat, Galen Andrew, Salman Avestimehr, Katharine Daly, Deepesh Data, Suhas Diggavi, Hubert Eichner, Advait Gadhikar, Zachary Garrett, Antonious M. Girgis, Filip Hanzely, Andrew Hard, Chaoyang He, Samuel Horvath, Zhouyuan Huo, Alex Ingerman, Martin Jaggi, Tara Javidi, Peter Kairouz, Satyen Kale, Sai Praneeth Karimireddy, Jakub Konecny, Sanmi Koyejo, Tian Li, Luyang Liu, Mehryar Mohri, Hang Qi, Sashank J. Reddi, Peter Richtarik, Karan Singhal, Virginia Smith, Mahdi Soltanolkotabi, Weikang Song, Ananda Theertha Suresh, Sebastian U. Stich, Ameeet Talwalkar, Hongyi Wang, Blake Woodworth, Shanshan Wu, Felix X. Yu, Honglin Yuan, Manzil Zaheer, Mi Zhang, Tong Zhang, Chunxiang Zheng, Chen Zhu, and Wennan Zhu. A Field Guide to Federated Optimization. *arXiv preprint arXiv:2107.06917*, 2021.
- Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American statistical association*, pages 63–69, 1965.
- Bin Yu. *Assouad, Fano, and Le Cam*. Springer New York, New York, NY, 1997. ISBN 978-1-4612-1880-7. doi: 10.1007/978-1-4612-1880-7\_29. URL [https://doi.org/10.1007/978-1-4612-1880-7\\_29](https://doi.org/10.1007/978-1-4612-1880-7_29).
- Razieh Nokhbeh Zaeem and K. Suzanne Barber. The effect of the gdpr on privacy policies: Recent progress and future promise. *ACM Trans. Manage. Inf. Syst.*, 12(1), December 2020. ISSN 2158-656X. doi: 10.1145/3389685. URL <https://doi.org/10.1145/3389685>.
- Puning Zhao, Lifeng Lai, Li Shen, Qingming Li, Jiafei Wu, and Zhe Liu. A Huber Loss Minimization Approach to Mean Estimation under User-level Differential Privacy. In *Advances in Neural Information Processing Systems*, 2024.



## S1 Proof of Theorem 1

**Theorem 1** (Lower bound). *Assume **H1-3**. Then, there exist  $c_1, c_2 > 0$ , independent of  $\alpha$ ,  $n$  and  $m$ , such that the following lower bound holds:*

$$R_{\alpha, n, M} \geq \max_{a \in \mathbb{N}} \frac{c_1 e^{-c_2 n \alpha^2 \mathbb{P}_{m \sim M}(m > a)^2}}{n \alpha^2 \mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}]^2 \vee 1} . \quad (10)$$

The positive constants  $c_1$  and  $c_2$  are explicitly given in Appendix **S1**.

The explicit values of the constants are  $c_1 = \frac{e^{-9}}{16}$  and  $c_2 = 24$ .

We start by the following result which relates the total variation of  $\nu_0$  and  $\nu_1$  to the total variation between  $\mu_0$  and  $\mu_1$ .

**Lemma S1** (Decomposition of TV distance by number of users). *Consider  $\nu_0 = \nu_{\mu_0}$  and  $\nu_1 = \nu_{\mu_1}$  where  $\nu_\mu$  is defined as in Equation (1). The following property holds:*

$$D_{TV}(\nu_0, \nu_1)^2 \leq \left( \sum_{m=1}^{\infty} M(m) \min\left(\sqrt{\frac{m}{2}} D_{KL}(\mu_0, \mu_1), 1\right) \right)^2 \quad (S1)$$

*Proof.* For the proof of Lemma **S1**, the following result will be used.

**Lemma S2** (TV and KL distance between  $\mu_0$  and  $\mu_1$ ). *Define  $\mu_0$  and  $\mu_1$  to be probability measures on  $\{-1, 1\}$  where  $\mu_0(\{-1\}) = (1 - \delta)/2$ , and  $\mu_0(\{1\}) = (1 + \delta)/2$ . It holds that*

$$D_{TV}(\mu_0, \mu_1) = \delta \quad (S2)$$

$$D_{KL}(\mu_0, \mu_1) \leq 3\delta^2 \text{ for } \delta \in [0, 1/2] \quad (S3)$$

*Proof of Lemma S2.* Let us start by showing (S2).

$$D_{TV}(\mu_0, \mu_1) = \frac{1}{2} (|\mu_0(\{1\}) - \mu_1(\{1\})| + |\mu_0(\{0\}) - \mu_1(\{0\})|) \quad (S4)$$

$$= \frac{1}{2} (|(1 - \delta)/2 - (1 + \delta)/2| + |(1 + \delta)/2 - (1 - \delta)/2|) \quad (S5)$$

$$= \delta \quad (S6)$$

Now let us prove (S3)

$$D_{KL}(\mu_0, \mu_1) = \mu_0(\{1\}) \cdot \ln\left(\frac{\mu_0(\{1\})}{\mu_1(\{1\})}\right) + \mu_0(\{0\}) \cdot \ln\left(\frac{\mu_0(\{0\})}{\mu_1(\{0\})}\right) \quad (S7)$$

$$= \delta \ln\left(\frac{1 + \delta}{1 - \delta}\right) \quad (S8)$$

$$\leq 3\delta^2 \text{ for } \delta \in [0, 1/2] \quad (S9)$$

where the last inequality comes since the function  $f(\delta) = \ln(\frac{1+\delta}{1-\delta}) - 3\delta$  is negative on  $[0, \frac{1}{2}]$ . Indeed,  $f(0) = 0$  and

$$f'(\delta) = \underbrace{\frac{1}{1+\delta}}_{\leq 1} + \underbrace{\frac{1}{1-\delta}}_{\leq 2} - 3 \leq 0.$$

□

First let us show that

$$D_{TV}(\nu_0, \nu_1) = \sum_{m=1}^{\infty} M(m) D_{TV}(\mu_0^i, \mu_1^i)$$

$$D_{TV}(\nu_0, \nu_1) = 2 \sup_{A \in \mathcal{X}} |\nu_0(A) - \nu_1(A)| \quad (\text{S10})$$

$$= 2 \sup_{A \in \mathcal{X}} \left| \nu_0 \left( \bigcup_{i=1}^{+\infty} A_i \right) - \nu_1 \left( \bigcup_{i=1}^{+\infty} A_i \right) \right| \quad (\text{Since } A = \bigcup_{i=1}^{+\infty} A_i, A_i \subset [-1, 1]^i \forall i \in \mathbb{N}^*) \quad (\text{S11})$$

$$= 2 \sup_{A \in \mathcal{X}} \left| \sum_{i=1}^{+\infty} (\nu_0(A_i \cap \{m=i\}) - \nu_1(A_i \cap \{m=i\})) \right| \quad (\text{S12})$$

$$= 2 \sup_{A \in \mathcal{X}} \left| \sum_{i=1}^{+\infty} p_i (\mu_0^i(A_i) - \mu_1^i(A_i)) \right| \quad (\text{S13})$$

$$\leq \sum_{i=1}^{+\infty} p_i \left( 2 \sup_{B \in [-1, 1]^i} (\mu_0^i(B) - \mu_1^i(B)) \right) \quad (\text{By triangular inequality}) \quad (\text{S14})$$

$$\leq \sum_{i=1}^{+\infty} p_i D_{TV}(\mu_0^i, \mu_1^i) \quad (\text{S15})$$

$$(\text{S16})$$

Then, notice that

$$D_{TV}(\nu_0, \nu_1)^2 = \left( \sum_{m=1}^{\infty} M(m) D_{TV}(\mu_0^m, \mu_1^m) \right)^2 \quad (\text{S17})$$

$$\stackrel{(1)}{\leq} \left( \sum_{m=1}^{\infty} M(m) \min(\sqrt{\frac{1}{2} D_{KL}(\mu_0^m, \mu_1^m)}, 1) \right)^2 \quad (\text{S18})$$

$$\stackrel{(2)}{=} \left( \sum_{m=1}^{\infty} M(m) \min(\sqrt{\frac{m}{2} D_{KL}(\mu_0, \mu_1)}, 1) \right)^2 \quad (\text{S19})$$

where inequality (1) is by Pinsker's inequality [Duchi \(2019\)](#) and  $D_{TV}(\mu_0^m, \mu_1^m) \leq 1$  and equality (2) is by the KL of product distributions.  $\square$

We now give the proof of Theorem 1. Define  $\mu_0$  and  $\mu_1$  to be probability measures on  $\{-1, 1\}$  where  $\mu_0(\{1\}) = (1 - \delta)/2$ , and  $\mu_1(\{1\}) = (1 + \delta)/2$  where the parameter  $\delta \in (0, \frac{1}{2})$  will be chosen later. Define  $\nu_0 = \nu_{\mu_0}$  and  $\nu_1 = \nu_{\mu_1}$  where  $\nu_{\mu}$  is defined as in Equation (1) and their privatized counterpart  $M_0^n, M_1^n$  defined for  $i \in \{0, 1\}$  as

$$M_i^n(S) = \int_{x \in \mathcal{X}} Q(S|x) d\nu_i(x).$$

Starting from Le Cam's bound [Duchi \(2019\)](#), we obtain the following results.

$$R_{n,\alpha,M} \geq \frac{\delta^2}{2} \{1 - D_{TV}(M_0^n, M_1^n)\} \quad (\text{Le Cam's bound}) \quad (\text{S20})$$

$$\geq \frac{\delta^2}{4} e^{-D_{KL}(M_0^n, M_1^n)} \quad (\text{Bretagnolle-Huber, Duchi (2019)}) \quad (\text{S21})$$

$$\geq \frac{\delta^2}{4} e^{-12n\alpha^2 D_{TV}(\nu_0, \nu_1)^2} \quad ((\text{Duchi et al., 2014, Th 1})). \quad (\text{S22})$$

We then upper bound  $D_{\text{TV}}(\nu_0, \nu_1)$  starting from Lemma S1.

$$D_{\text{TV}}(\nu_0, \nu_1)^2 \leq \left( \sum_{m=1}^{+\infty} M(m) \min\left(\sqrt{\frac{m}{2}} D_{KL}(\mu_0, \mu_1), 1\right) \right)^2 \quad (\text{By Lemma S1}) \quad (\text{S23})$$

$$\leq \left( \sum_{m=1}^{+\infty} M(m) \min\left(\sqrt{\frac{m}{2}} 3\delta^2, 1\right) \right)^2. \quad (\text{By Lemma S2}) \quad (\text{S24})$$

Then for any  $a \in \mathbb{N}^*$  it holds that

$$D_{\text{TV}}(\nu_0, \nu_1)^2 \leq \left( \sum_{m=1}^a M(m) \min\left(\sqrt{\frac{m}{2}} 3\delta^2, 1\right) + \sum_{m=a+1}^{+\infty} M(m) \min\left(\sqrt{\frac{m}{2}} 3\delta^2, 1\right) \right)^2 \quad (\text{S25})$$

$$\leq \left( \sum_{m=1}^a M(m) \min\left(\sqrt{\frac{m}{2}} 3\delta^2, 1\right) + \sum_{m=a+1}^{+\infty} M(m) \right)^2 \quad (\text{S26})$$

$$\leq \left( \sqrt{\frac{3\delta^2}{2}} \mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}] + \mathbb{P}_{m \sim M}(m > a) \right)^2 \quad (\text{S27})$$

$$\leq 3\delta^2 \mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}]^2 + 2\mathbb{P}_{m \sim M}(m > a)^2 \quad (\text{S28})$$

$$(\text{S29})$$

Then choosing  $\delta^2 = \frac{1}{4 \max(n\alpha^2 \mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}]^2, 1)} \leq \frac{1}{4}$  we obtain the bound

$$\forall a \in \mathbb{N}^* \quad (\text{S30})$$

$$R_{n,\alpha,M} \geq \frac{1}{16 \max(\mathbb{E}_{m \sim M}[\sqrt{m} \mathbb{1}\{m \leq a\}]^2 n\alpha^2, 1)} \exp(-9) \exp(-24n\alpha^2 \mathbb{P}_{m \sim M}(m > a)^2), \quad (\text{S31})$$

which in turn gives the bound stated in the theorem.

## S2 Proof of Theorem 2

**Theorem 2** (Upper bound). *Assume H1-3. Let  $c_3, c_4, c_5 > 0$  be universal constants, independent of  $n, \alpha$  and  $M$ . Consider the function  $\phi : a \in \mathbb{N}^* \mapsto c_5 / (n\alpha^2) \log[c_4(an\alpha^2 \vee 1) / \log(c_4(an\alpha^2 \vee 1))]$  and set  $\tilde{m} = \arg\max_{a \in \mathbb{N}^*} \{\mathbb{P}_{m \sim M}(m \geq a)^2 \geq \phi(a) \wedge 1\}$ . Then, we have the following lower bound.*

$$R_{\alpha,n,M} \leq \frac{c_3 \ln(c_4(\sqrt{\tilde{m}n\alpha^2} \vee 1))}{n\alpha^2 (\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m}))^2} \wedge 4. \quad (11)$$

The positive constants  $c_3, c_4$  and  $c_5$  are explicitly given in Appendix S2.

*Proof.* We will show that  $c_3 = 1570, c_4 = 8, c_5 = 868.5$ . The fact that  $\mathbb{E}[|\hat{\theta} - \theta|^2] \leq 4$  comes from  $\hat{\theta}, \theta \in [-1, 1]$ . To get the other term in the bound, we start by writing the following bias-variance decomposition.

$$\mathbb{E}[|\hat{\theta} - \theta|^2] = \mathbb{E}\left[\mathbb{E}\left\{|\hat{\theta} - \theta|^2 \mid \hat{j}\right\}\right] \quad (\text{S32})$$

$$= \mathbb{E}\left[\left\{\mathbb{E}\left(\hat{\theta} \mid \hat{j}\right) - \theta\right\}^2 + \text{Var}\left(\hat{\theta} \mid \hat{j}\right)\right]. \quad (\text{S33})$$

where  $\hat{\theta}$  is defined Equation (9). Let  $m$  be a random variable sampled from  $M$ . It holds that

$$\hat{\theta} = \frac{\sqrt{\tilde{m}} \frac{2}{n} \sum_{u=n/2+1}^n \hat{\theta}^{(u)}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_j, \quad (\text{S34})$$

First, let us focus on the biased term. We can write that:

$$\mathbb{E}(\hat{\theta} \mid \hat{j}) = \mathbb{E}\left(\frac{\sqrt{\tilde{m}} \frac{2}{n} \sum_{u=n/2+1}^n \hat{\theta}^{(u)}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}} \mid \hat{j}\right) \quad (\text{S35})$$

$$= \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}\left(\frac{2}{n} \sum_{u=n/2+1}^n \hat{\theta}^{(u)} \mid \hat{j}\right) - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}} \quad (\text{S36})$$

$$= \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}(\hat{\theta}^{(n)} \mid \hat{j}) - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}} \quad (\text{S37})$$

$$\stackrel{(1)}{=} \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}\left(\Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) \mid \hat{j}\right) - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}} \\ \stackrel{(2)}{=} \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}\left(\Pi_{\hat{j}}(\bar{X}_{m_n}^{(n)}) \mathbb{1}(m_n \geq \tilde{m}) + \sum_{i=1}^{\tilde{m}-1} \Pi_{\hat{j}}(\bar{X}_{m_n}^{(n)}) \mathbb{1}(m_n = i) \mid \hat{j}\right) \\ - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}}$$

where equality (1) comes from the definition of  $\hat{\theta}^{(n)}$  in (8) and equality (2) comes from the definition of  $\hat{X}_{\hat{j}}^{(n)}$  in (3).

Then, we can write  $\theta = \mathbb{E}_{X \sim \mu}[X]$  in function of  $\bar{X}_{m_n}^{(n)} \mathbb{1}\{m_n \geq \tilde{m}\}$  and  $(\bar{X}_i^{(n)})_{i=1}^{\tilde{m}}$ .

**Lemma S3.** *Let  $\theta = \mathbb{E}_{X \sim \mu}[X]$ ,  $m \sim M$ . It holds that*

$$\theta = \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}(\bar{X}_{m_n}^{(n)} \mathbb{1}(m_n \geq \tilde{m})) + \frac{1}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \sum_{i=1}^{\tilde{m}-1} \mathbb{E}(\bar{X}_i^{(n)} \sqrt{i} \mathbb{1}(m_n = i)). \quad (\text{S38})$$

*Proof.* We can write

$$\underbrace{\frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}(\bar{X}_{m_n}^{(n)} \mathbb{1}(m_n \geq \tilde{m})) + \frac{1}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \sum_{i=1}^{\tilde{m}-1} \mathbb{E}(\bar{X}_{m_n}^{(n)} \sqrt{i} \mathbb{1}(m_n = i))}_{(i)} \quad (\text{S39})$$

$$= \sum_{i=\tilde{m}}^{\infty} \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}(\bar{X}_{m_n}^{(n)} \mathbb{1}\{m_n = i\}) + \frac{1}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \sum_{i=1}^{\tilde{m}-1} \sqrt{i} \mathbb{E}(\bar{X}_{m_n}^{(n)} \mathbb{1}(m_n = i)). \quad (\text{S40})$$

Then notice that

$$\mathbb{E}(\bar{X}_{m_n}^{(n)} \mathbb{1}\{m_n = i\}) = \mathbb{E}[\underbrace{\mathbb{E}[\bar{X}_i^{(n)} \mid m_n = i]}_{=\theta} \mathbb{1}\{m_n = i\}] \quad (\text{S41})$$

$$= \theta M(i), \quad (\text{S42})$$

so that

$$(i) = \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \sum_{i=\tilde{m}}^{+\infty} \theta M(i) + \frac{1}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \sum_{i=1}^{\tilde{m}-1} \sqrt{i} M(i) \theta \quad (\text{S43})$$

$$= \theta \quad (\text{S44})$$

□



Then using Lemma S3, we get

$$\mathbb{E}(\hat{\theta} | \hat{j}) - \theta \quad (\text{S45})$$

$$= \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}(\Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) | \hat{j}) - \frac{\mathbb{E}((\sqrt{\tilde{m}} - \sqrt{m})\mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}} \quad (\text{S46})$$

$$- \left( \frac{\sqrt{\tilde{m}}}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}(\bar{X}_{m_n}^{(n)} \mathbb{1}(m_n \geq \tilde{m})) + \frac{1}{\mathbb{E}(\sqrt{m} \wedge \tilde{m})} \sum_{i=1}^{\tilde{m}-1} \mathbb{E}(\bar{X}_{m_n}^{(n)} \sqrt{i} \mathbb{1}(m_n = i)) \right). \quad (\text{S47})$$

and using the independence between  $(\bar{X}_{m_n}^{(n)}, m_n)$  and  $\hat{j}$ , we obtain

$$\begin{aligned} \mathbb{E}(\hat{\theta} | \hat{j}) - \theta &= \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}\left(\left(\Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \bar{X}_{m_n}^{(n)}\right) \mathbb{1}(m_n \geq \tilde{m}) | \hat{j}\right) \\ &+ \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} \sum_{i=1}^{\tilde{m}-1} \mathbb{E}\left(\left(\Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_{m_n}^{(n)}\right) \mathbb{1}(m_n = i) | \hat{j}\right) \\ &- \frac{\mathbb{E}_{m \sim M}((\sqrt{\tilde{m}} - \sqrt{m})\mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}}. \end{aligned} \quad (\text{S48})$$

We now introduce two events that will help us further upper bound the bias. Given  $l \in \mathbb{N}^*$  the index of the bin  $I_l$  such that  $\theta \in I_l$ , we define the event that  $\hat{j}$  gives the correct bin up to an error of 2:

$$\mathcal{A} = \{|\hat{j} - l| \leq 2\}. \quad (\text{S49})$$

Note that under  $\mathcal{A}$ , since  $s_{\hat{j}} \in I_{\hat{j}}$ , it follows that  $|s_{\hat{j}} - \theta| \leq 5\tau$  since  $\forall j \in \frac{1}{\tau}, |I_j| = 2\tau$ .

The bias can be decomposed depending on whether the event  $\mathcal{A}$  holds:

$$\mathbb{E}[(\mathbb{E}[\hat{\theta} | \hat{j}] - \theta)^2] = \mathbb{E}[(\mathbb{E}[\hat{\theta} | \hat{j}] - \theta)^2 \mathbb{1}\{\mathcal{A}\}] + \mathbb{E}[(\mathbb{E}[\hat{\theta} | \hat{j}] - \theta)^2 \mathbb{1}\{\bar{\mathcal{A}}\}] \quad (\text{S50})$$

$$\leq \mathbb{E}[(\mathbb{E}[\hat{\theta} | \hat{j}] - \theta)^2 \mathbb{1}\{\mathcal{A}\}] + 4\mathbb{P}(\bar{\mathcal{A}}) \quad (\text{S51})$$

$$= \mathbb{E}[(\mathbb{E}[\hat{\theta} | \hat{j}] - \theta)^2 \mathbb{1}\{\mathcal{A}\}] + 4\mathbb{P}(\bar{\mathcal{A}}) \quad (\text{S52})$$

We first focus on upper bounding  $\mathbb{P}(\bar{\mathcal{A}})$ .

**Lemma S4** (Hoeffding bound [Hoeffding \(1963\)](#)). *The random variables  $(X_t^{(u)})_{t=1}^i$  have value in  $[-1, 1]$  and given  $m_u = i$  are independent and identically distributed with mean  $\theta$ . It holds that*

$$\mathbb{P}\left\{\left|\bar{X}_i^{(u)} - \theta\right| \geq \tau \mid m_u = i\right\} \leq 2 \exp(-i\tau^2/2) \quad (\text{S53})$$

**Lemma S5** (Upper bound on  $\mathbb{P}(\bar{\mathcal{A}})$ ). *It holds that*

$$\mathbb{P}(\bar{\mathcal{A}}) \leq \frac{1}{\tau} \exp\left(-n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579\right). \quad (\text{S54})$$

*Proof.* Call  $l$  the optimal index i.e. the index such that  $\theta \in I_l$ .

User  $u$  sets  $V_l^{(u)} = 0$  because either it does not have enough samples ( $m_u \leq \tilde{m}$ ) or because it has enough samples but its mean estimate is far from  $\theta$ . Formally,

$$\{V_l^{(u)} = 0\} \subseteq \{m_u \geq \tilde{m}\} \cap \left\{\left|\bar{X}^{(u)} - \theta\right| \geq \tau\right\} \cup \{m_u \leq \tilde{m}\}. \quad (\text{S55})$$

It follows that,

$$\mathbb{P}(V_l^{(u)} = 0) \leq \sum_{i=\tilde{m}}^{+\infty} M(i) \mathbb{P}\left\{\left|\bar{X}^{(u)} - \theta\right| \geq \tau \mid m_u = i\right\} + 1 - \mathbb{P}(m \geq \tilde{m}) \quad (\text{S56})$$

$$\leq \sum_{i=\tilde{m}}^{+\infty} M(i) 2 \exp(-i\tau^2/2) + 1 - \mathbb{P}(m \geq \tilde{m}) \quad (\text{By Lemma S4}) \quad (\text{S57})$$

$$\leq 2\mathbb{P}(m_u \geq \tilde{m}) \exp(-\tilde{m}\tau^2/2) + 1 - \mathbb{P}(m_u \geq \tilde{m}). \quad (\text{S58})$$

Let  $k \in [2/\tau] \setminus \{l-2, l-1, l, l+1, l+2\}$  a sub-optimal index. In order to have  $V_k^{(u)} = 1$  for a suboptimal index, the user  $u$  must have enough samples but have a mean estimate far from  $\theta$ . Formally,

$$\{V_k^{(u)} = 1\} \subseteq \{m_u \geq \tilde{m}\} \cap \left\{ \left| \bar{X}^{(u)} - \theta \right| \geq \tau \right\}, \quad (\text{S59})$$

and therefore,

$$\mathbb{P}\left(V_k^{(i)} = 1\right) \leq 2 \exp(-\tilde{m}\tau^2/2) \mathbb{P}(m \geq \tilde{m}). \quad (\text{S60})$$

Then, the following equation relates for any  $j \in [\frac{1}{\tau}]$ ,  $V_j^{(u)}$  and its privatized version  $\tilde{V}_j^{(u)}$ . Writing for any  $x > 0$ ,  $\pi_x = e^x / (1 + e^x)$ , we have that for any  $j \in [\frac{1}{\tau}]$ ,

$$p_j \triangleq \mathbb{P}\left(\tilde{V}_j^{(i)} = 1\right) \quad (\text{S61})$$

$$= \pi_{\alpha/6} \mathbb{P}\left(V_j^{(i)} = 1\right) + (1 - \pi_{\alpha/6}) \mathbb{P}\left(V_j^{(i)} = 0\right) \quad (\text{S62})$$

$$= \pi_{\alpha/6} \mathbb{P}\left(V_j^{(i)} = 1\right) + (1 - \pi_{\alpha/6})(1 - \mathbb{P}\left(V_j^{(i)} = 1\right)) \quad (\text{S63})$$

$$= (2\pi_{\alpha/6} - 1) \mathbb{P}\left(V_j^{(i)} = 1\right) + (1 - \pi_{\alpha/6}).$$

Hence,

$$p_l = \mathbb{P}\left(\tilde{V}_l^{(i)} = 1\right) \quad (\text{S64})$$

$$= (2\pi_{\alpha/6} - 1) \left(1 - \mathbb{P}\left(V_l^{(i)} = 0\right)\right) + (1 - \pi_{\alpha/6}) \quad (\text{S65})$$

$$\geq (2\pi_{\alpha/6} - 1)(\mathbb{P}(m_u \geq \tilde{m}) - 2\mathbb{P}(m_u \geq \tilde{m}) \exp(-\tilde{m}\tau^2/2)) + (1 - \pi_{\alpha/6}),$$

and similarly,

$$p_k \leq 2(2\pi_{\alpha/6} - 1) \exp(-\tilde{m}\tau^2/2) \mathbb{P}(m \geq \tilde{m}) + (1 - \pi_{\alpha/6}).$$

We therefore get

$$\begin{aligned} p_l - p_k &\geq -4(2\pi_{\alpha/6} - 1) \mathbb{P}(m_u \geq \tilde{m}) \exp(-\tilde{m}\tau^2/2) + \mathbb{P}(m_u \geq \tilde{m}) (2\pi_{\alpha/6} - 1) \\ &= (2\pi_{\alpha/6} - 1) \mathbb{P}(m_u \geq \tilde{m}) (1 - 4 \exp(-\tilde{m}\tau^2/2)) \end{aligned} \quad (\text{S66})$$

It holds that  $(1 - 4 \exp(-\tilde{m}\tau^2/2)) \geq 1/2$ , since by definition of  $\tau$  in Equation (4),  $\tau \geq \sqrt{\frac{2 \ln(8)}{\tilde{m}}}$ .

Hence, we get :

$$p_l - p_k \geq (\pi_{\alpha/6} - 1/2) \mathbb{P}(m_u \geq \tilde{m}) \quad (\text{S67})$$

Letting  $Z_j = \sum_{i=1}^{n/2} \tilde{V}_j^{(i)}$  denote the total privatised votes for the  $j$ -th sub-interval.

It holds that:

$$\mathcal{A} = \{|j - l| \leq 2\} \quad (\text{S68})$$

$$= \bigcap_{k \in [\frac{1}{\tau}] : |l - k| > 2} \{Z_l > Z_k\} \quad (\text{S69})$$

We note that

$$Z_l - Z_k = \sum_{i=1}^{n/2} \left( \tilde{V}_l^{(u)} - \tilde{V}_k^{(u)} \right)$$

where for each  $u \in [n/2]$ , it holds that  $\tilde{V}_l^{(u)} - \tilde{V}_k^{(u)} \in \{-1, 0, 1\}$  and that  $\mathbb{E} \left[ \tilde{V}_l^{(u)} - \tilde{V}_k^{(u)} \right] = p_l - p_k$ . Then

$$\begin{aligned} \mathbb{P}(Z_l \leq Z_k) &= \mathbb{P} \left( \sum_{u=1}^{n/2} \left( \tilde{V}_l^{(u)} - \tilde{V}_k^{(u)} \right) / n \leq 0 \right) \\ &\leq \exp(-n/4(p_l - p_k)^2) \end{aligned} \quad \text{(By Hoeffding inequality)} \quad (\text{S70})$$

$$\leq \exp \left( -n/4 \left( \pi_{\alpha/6} - 1/2 \right)^2 \mathbb{P}(m_u \geq \tilde{m})^2 \right) \quad \text{(By Equation (S67))} \quad (\text{S71})$$

Lemma S6 that we prove next shows that  $(\pi_{\alpha/6} - 1/2)^2 \geq \alpha^2/579$  so that we get

$$\mathbb{P}(Z_l \leq Z_j) \leq \exp \left( -n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579 \right).$$

**Lemma S6.** *For  $\alpha \in [0, 1]$ , it holds that*

$$(1/2 - \pi_{\alpha/6})^2 \geq \alpha^2/579 \quad (\text{S72})$$

*Proof.* Consider the function

$$f : x \in [0, 1] \mapsto \frac{x^2}{\left( 1/2 - \frac{e^{x/6}}{1+e^{x/6}} \right)^2} \quad (\text{S73})$$

Since,

$$f(x) = \frac{x^2}{\left( 1/2 - \frac{\exp(x/6)}{1+\exp(x/6)} \right)^2} \quad (\text{S74})$$

$$\sim_{x \rightarrow 0} \frac{x^2}{\left( 1/2 - \frac{1+x/6}{2+x/6} \right)^2} \quad (\text{S75})$$

$$\sim_{x \rightarrow 0} \frac{x^2}{\left( 1/2 - \frac{1+x/6}{2} (1 - x/12) \right)^2} \quad (\text{S76})$$

$$\sim_{x \rightarrow 0} \frac{x^2}{\left( 1/2 - (1/2 + x/24) \right)^2} \quad (\text{S77})$$

$$\sim_{x \rightarrow 0} 24^2. \quad (\text{S78})$$

And,  $f(x) > 0 \forall x \in (0, 1]$  We can define :

$$g : x \in [0, 1] \mapsto \ln(f(x)) \quad (\text{S79})$$

$\forall x \in [0, 1]$

$$g'(x) = \frac{2}{x} + \frac{1}{\frac{1}{2} - \frac{\exp(x/6)}{1+\exp(x/6)}} \cdot \frac{\frac{1}{6} \exp(x/6)}{(1 + \exp(x/6))^2} \quad (\text{S80})$$

$$= \frac{2}{x} + \frac{2(1 + \exp(x/6))}{(1 - \exp(x/6))} \cdot \frac{\frac{1}{6} \exp(x/6)}{(1 + \exp(x/6))^2} \quad (\text{S81})$$

$$= \frac{2}{x} + \frac{2}{(1 - \exp(x/6))} \cdot \frac{\frac{1}{6} \exp(x/6)}{(1 + \exp(x/6))} \quad (\text{S82})$$

$$= \frac{2}{x} + \frac{1/3 \exp(x/6)}{1 - \exp(x/3)} \quad (\text{S83})$$

$$\geq \frac{2}{x} - \frac{1/3 \exp(1/6)}{\exp(x/3) - 1} \quad \exp(x/6) \leq \exp(1/6) \quad \forall x \in [0, 1] \quad (\text{S84})$$

$$\geq \frac{2}{x} - \frac{1/3 \exp(1/6)}{x/3} \quad \exp(x/3) - 1 \geq \frac{x}{3} \quad \forall x \in \mathbb{R} \quad (\text{S85})$$

$$= \frac{2 - \exp(1/6)}{x} \quad (\text{S86})$$

$$\geq 0 \quad (\text{S87})$$

Hence,  $g$  is non decreasing on  $[0, 1]$ , and so is  $f$ . We get

$$\forall x \in [0, 1], f(x) \leq f(1) \leq 579. \quad (\text{S88})$$

□

We can then finish the proof of Lemma S5 by writing

$$\begin{aligned} \mathbb{P}(\overline{\mathcal{A}}) &\leq \sum_{j \in \lceil 1/\tau \rceil : |l-j| > 2} \mathbb{P}(Z_l \leq Z_j) \quad (\text{By a union bound}) \\ &\leq \frac{1}{\tau} \exp\left(-n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579\right) \end{aligned} \quad (\text{S89})$$

□

We then focus on  $|\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta| \mathbb{1}\{A\}$ . From the triangle inequality, the following equation holds.

$$|\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta| \mathbb{1}\{A\} \quad (\text{S90})$$

$$\leq \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} \underbrace{\mathbb{E}\left(\left|\Pi_{\hat{j}}\left(\hat{X}_{\hat{j}}^{(n)}\right) - \overline{X}_{m_n}^{(n)}\right| \mathbb{1}(m_n \geq \tilde{m}) \mathbb{1}\{A\} \mid \hat{j}\right)}_{(I)} \quad (\text{S91})$$

$$+ \left| \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} \mathbb{E}\left(\sum_{i=1}^{\tilde{m}-1} \left(\Pi_{\hat{j}}\left(\hat{X}_{\hat{j}}^{(n)}\right) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \overline{X}_{m_n}^{(n)}\right) \mathbb{1}(m_n = i) \mathbb{1}\{A\} \mid \hat{j}\right) \right| \quad (\text{S92})$$

$$- \frac{\mathbb{E}_{m \sim M}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}} \mathbb{1}\{A\} \Big|. \quad (\text{By Equation (S48)}) \quad (\text{S93})$$

We therefore seek to upper bound the term (I).

**Lemma S7** (Majoration of (I)).

$$\mathbb{E}\left(\left|\Pi_{\hat{j}}\left(\hat{X}_{\hat{j}}^{(n)}\right) - \overline{X}_{m_n}^{(n)}\right| \mathbb{1}(m_n \geq \tilde{m}) \mathbb{1}\{A\} \mid \hat{j}\right) \leq 4 \exp(-\tilde{m}\tau^2/2) \leq \frac{1}{4\sqrt{\tilde{m}n\alpha^2}} \quad (\text{S94})$$



*Proof.* The proof follows by a similar argument as in [Duchi et al. \(2018\)](#).

We have

$$\mathbb{E} \left( \left| \Pi_{\hat{j}} \left( \hat{X}_{\hat{j}}^{(n)} \right) - \bar{X}_{m_n}^{(n)} \right| \mathbb{1}(m_n \geq \tilde{m}) \mathbb{1}\{\mathcal{A}\} \mid \hat{j} \right) \quad (\text{S95})$$

$$= \mathbb{E} \left[ \left| \bar{X}_{m_n}^{(n)} - \Pi_{\hat{j}} \left( \bar{X}_{m_n}^{(n)} \right) \right| \mathbb{1}\{m_n \geq \tilde{m}\} \mathbb{1}\{\mathcal{A}\} \mid \hat{j} \right] \quad (\text{S96})$$

$$= \mathbb{E} \left[ \left| \bar{X}_{m_n}^{(n)} - \tilde{U}_{\hat{j}} \right| \mathbb{1} \left\{ \bar{X}_{m_n}^{(n)} > \tilde{U}_{\hat{j}} \right\} \mathbb{1}\{m_n \geq \tilde{m}\} \mathbb{1}\{\mathcal{A}\} \mid \hat{j} \right] \\ + \mathbb{E} \left[ \left| \bar{X}_{m_n}^{(n)} - \tilde{L}_{\hat{j}} \right| \mathbb{1} \left\{ \bar{X}_{m_n}^{(n)} < \tilde{L}_{\hat{j}} \right\} \mathbb{1}\{m_n \geq \tilde{m}\} \mathbb{1}\{\mathcal{A}\} \mid \hat{j} \right] \quad (\text{S97})$$

$$\leq 2\mathbb{P} \left( \left\{ \bar{X}_{m_n}^{(n)} > \tilde{U}_{\hat{j}} \right\} \cap \{m_n \geq \tilde{m}\} \cap \mathcal{A} \mid \hat{j} \right) \\ + 2\mathbb{P} \left( \left\{ \bar{X}_{m_n}^{(n)} < \tilde{L}_{\hat{j}} \right\} \cap \{m_n \geq \tilde{m}\} \cap \mathcal{A} \mid \hat{j} \right), \quad (\text{S98})$$

where in the last inequality uses that  $|\bar{X}_{m_n}^{(n)} - \tilde{L}_{\hat{j}}| \leq 2$  and  $|\bar{X}_{m_n}^{(n)} - \tilde{U}_{\hat{j}}| \leq 2$  which follows since  $\bar{X}_{m_n}^{(n)}, \tilde{L}_{\hat{j}}, \tilde{U}_{\hat{j}} \in [-1, 1]$ . Then, we upper bound  $\mathbb{P} \left( \left\{ \bar{X}_{m_n}^{(n)} > \tilde{U}_{\hat{j}} \right\} \cap \{m_n \geq \tilde{m}\} \cap \mathcal{A} \mid \hat{j} \right)$ .

$$\mathbb{P} \left( \left\{ \bar{X}_{m_n}^{(n)} > \tilde{U}_{\hat{j}} \right\} \cap \{m_n \geq \tilde{m}\} \cap \mathcal{A} \mid \hat{j} \right) \quad (\text{S99})$$

$$= \mathbb{P} \left( \bigcup_{i=\tilde{m}}^{+\infty} \left\{ \bar{X}_i^{(n)} > \tilde{U}_{\hat{j}} \right\} \cap \{m_n = i\} \cap \mathcal{A} \mid \hat{j} \right) \quad (\text{S100})$$

$$= \sum_{i=\tilde{m}}^{+\infty} \mathbb{P} \left( \left\{ \bar{X}_i^{(n)} > \tilde{U}_{\hat{j}} \right\} \mid \{m_n = i\}, \hat{j} \right) M(i) \mathbb{P}(\mathcal{A}) \quad (\text{S101})$$

$$\leq \sum_{i=\tilde{m}}^{+\infty} \mathbb{P} \left( \left\{ \bar{X}_i^{(n)} > \tilde{U}_{\hat{j}} \right\} \mid \{m_n = i\}, \hat{j} \right) M(i) \quad (\text{S102})$$

$$= \sum_{i=\tilde{m}}^{+\infty} \mathbb{P} \left( \left\{ \bar{X}_i^{(n)} > \tilde{U}_{\hat{j}} \right\} \mid \{m_n = i\} \right) M(i) \quad (\text{Since } \hat{j} \text{ independent of } \bar{X}_i^{(n)} \text{ given } m_n = i) \quad (\text{S103})$$

$$\leq \sum_{i=\tilde{m}}^{+\infty} \mathbb{P} \left( \left\{ \bar{X}_i^{(n)} > \theta + \tau \right\} \mid \{m_n = i\} \right) M(i) \quad (\text{S104})$$

$$\leq \sum_{i=\tilde{m}}^{+\infty} \exp(-i(\tau)^2/2) M(i) \quad (\text{Hoeffding inequality}) \quad (\text{S105})$$

$$\leq \exp(-\tilde{m}\tau^2/2) \quad (\text{S106})$$

By a symmetric reasoning, we show that

$$\mathbb{P} \left( \left\{ \bar{X}_{m_n}^{(n)} < \tilde{L}_{\hat{j}} \right\} \cap \{m_n \geq \tilde{m}\} \cap \mathcal{A} \mid \hat{j} \right) \leq \exp(-\tilde{m}\tau^2/2).$$

Together the two bounds yield the result.  $\square$

We now introduce, for any  $a \in [\tilde{m} - 1]$  the events  $B_a$  that describe how  $\bar{X}_a^{(n)}$  concentrates around  $\theta$ :

$$\forall a \in [\tilde{m}], B_a = \left\{ \left| \bar{X}_a^{(n)} - \theta \right| \leq \frac{\sqrt{2 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}}{\sqrt{a}} \right\}. \quad (\text{S107})$$

By using Lemma S7 and incorporating the events  $(B_a)_{a \in [\tilde{m}-1]}$ , the upper bound on  $|\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta| \mathbb{1}\{\mathcal{A}\}$  becomes:

$$|\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta| \mathbb{1}\{\mathcal{A}\} \quad (\text{S108})$$

$$\leq \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})} 4 \exp(-\tilde{m}\tau^2/2) \quad (\text{S109})$$

$$+ \left| \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})} \mathbb{E} \left( \left[ \sum_{i=1}^{\tilde{m}-1} \left( \Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_{m_n}^{(n)} \right) \mathbb{1}(m_n = i) - \frac{\mathbb{E}_{m \sim M}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbb{1}(m \leq \tilde{m}))}{\sqrt{\tilde{m}}} s_{\hat{j}} \right] \mathbb{1}\{\mathcal{A}\} \mid \hat{j} \right) \right| \quad (\text{S110})$$

$$= \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})} 4 \exp(-m\tau^2/2) \quad (\text{S111})$$

$$+ \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})} \underbrace{\left| \mathbb{E} \left( \sum_{i=1}^{\tilde{m}-1} \left( \Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_{m_n}^{(n)} - \frac{(\sqrt{\tilde{m}} - \sqrt{i})}{\sqrt{\tilde{m}}} s_{\hat{j}} \right) \mathbb{1}\{\mathcal{A}, \mathcal{B}_i\} \mathbb{1}(m_n = i) \mid \hat{j} \right) \right|}_{(II)} \quad (\text{S112})$$

$$+ \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})} \underbrace{\left| \mathbb{E} \left( \sum_{i=1}^{\tilde{m}-1} \left( \Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_{m_n}^{(n)} - \frac{(\sqrt{\tilde{m}} - \sqrt{i})}{\sqrt{\tilde{m}}} s_{\hat{j}} \right) \mathbb{1}\{\mathcal{A}, \bar{\mathcal{B}}_i\} \mathbb{1}(m_n = i) \mid \hat{j} \right) \right|}_{(III)} \quad (\text{S113})$$

Let us start by upper bounding (III).

**Lemma S8** (Majoration of (III)).

$$(III) \leq \frac{1}{2\sqrt{\tilde{m}n\alpha^2}}$$

*Proof.* Since  $\bar{X}_{m_n}^{(n)} \in [-1, 1]$ ,  $\Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) \in [-1, 1]$  and  $s_{\hat{j}} \in [-1, 1]$ , we have that

$$\Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_{m_n}^{(n)} - \frac{(\sqrt{\tilde{m}} - \sqrt{i})}{\sqrt{\tilde{m}}} s_{\hat{j}} \in [-2, 2]$$

and therefore

$$(III) = \left| \mathbb{E} \left( \sum_{i=1}^{\tilde{m}-1} \left( \Pi_{\hat{j}}(\hat{X}_{\hat{j}}^{(n)}) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_{m_n}^{(n)} - \frac{(\sqrt{\tilde{m}} - \sqrt{i})}{\sqrt{\tilde{m}}} s_{\hat{j}} \right) \mathbb{1}\{\mathcal{A}, \bar{\mathcal{B}}_i\} \mathbb{1}(m_n = i) \mid \hat{j} \right) \right| \quad (\text{S114})$$

$$\leq 2 \sum_{i=1}^{\tilde{m}} \mathbb{P}(\bar{\mathcal{B}}_i, m_n = i \mid \hat{j}) \quad (\text{S115})$$

$$= 2 \sum_{i=1}^{\tilde{m}} \mathbb{P}(\bar{\mathcal{B}}_i, m_n = i) \quad (\text{Since } \hat{j} \text{ and } (m_n, \mathcal{B}_i) \text{ are independent}) \quad (\text{S116})$$

$$= 2 \sum_{i=1}^{\tilde{m}} \mathbb{P}(\bar{\mathcal{B}}_i \mid m_n = i) M(i) \quad (\text{S117})$$

$$= 2 \sum_{i=1}^{\tilde{m}} \mathbb{P}(|\bar{X}_i^{(n)} - \theta| > \frac{\sqrt{2 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}}{\sqrt{a}} \mid m_n = i) M(i) \quad (\text{S118})$$

$$\leq \frac{1}{2\sqrt{\tilde{m}n\alpha^2}} \sum_{i=1}^{\tilde{m}} M(i) \quad (\text{By Hoeffding bound}) \quad (\text{S119})$$

$$\leq \frac{1}{2\sqrt{\tilde{m}n\alpha^2}} \quad (\text{S120})$$

□

Then, we focus on (II). Under  $\mathcal{A}$  and  $\mathcal{B}_a$ , Lemma S9 upper bounds the distance between  $\hat{X}_j^{(n)}$  and  $s_j$ .

**Lemma S9** (Concentration of users with few samples).  $\forall a \leq \tilde{m}$ , it holds that

$$\mathcal{A} \cap \mathcal{B}_a \implies \left| \hat{X}_j^{(n)} - s_j \right| \mathbb{1}\{m_n = a\} \leq 6\tau. \quad (\text{S121})$$

*Proof.* Fix  $a \leq \tilde{m}$ . Under  $\mathcal{A}$  and  $\mathcal{B}_a$ , it holds that

$$\left| \hat{X}_j^{(m)} - s_j \right| \mathbb{1}\{m_n = a\} = \left| \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \left( \bar{X}_a^{(n)} + \left( \frac{\sqrt{\tilde{m}}}{\sqrt{a}} - 1 \right) s_j \right) - s_j \right| \mathbb{1}\{m_n = a\} \quad (\text{S122})$$

$$\leq \left| \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \left( \bar{X}_a^{(n)} + \left( \frac{\sqrt{\tilde{m}}}{\sqrt{a}} - 1 \right) s_j \right) - s_j \right| \quad (\text{S123})$$

$$= \left| \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \left( \bar{X}_a^{(n)} - s_j \right) \right| \quad (\text{S124})$$

$$\leq \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \left( \left| \bar{X}_a^{(n)} - \theta \right| + \left| \theta - s_j \right| \right) \quad (\text{S125})$$

$$\leq \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \left( \frac{\sqrt{2 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}}{\sqrt{a}} + 5\tau \right) \quad \text{Valid under } \mathcal{B}_a \cap \mathcal{A} \quad (\text{S126})$$

$$\leq 6\tau \quad (\text{S127})$$

□

From Lemma S9 we know that  $\hat{X}_j^{(n)}$  is at distance at most of  $6\tau$  of  $s_j$ , the center of  $I_j$ . Furthermore,  $\hat{X}_j^{(m)} \in [-1, 1]$ . Hence, as  $[\tilde{L}_j, \tilde{U}_j] = [-1 \vee (s_j - 6\tau), 1 \wedge (s_j + 6\tau)]$  we deduce that  $\hat{X}_j^{(n)} \in [\tilde{L}_j, \tilde{U}_j]$ . From the definition of  $\Pi_j$  which is the projection operator on  $[\tilde{L}_j, \tilde{U}_j]$ , it holds that

$$\begin{aligned} \forall a < \tilde{m}, \Pi_j(\hat{X}_j^{(n)}) \mathbb{1}\{m_n = a, \mathcal{A}, \mathcal{B}_a\} &= \hat{X}_j^{(n)} \mathbb{1}\{m_n = a, \mathcal{A}, \mathcal{B}_a\} \\ &= \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \left( \bar{X}_a^{(n)} + \left( \frac{\sqrt{\tilde{m}}}{\sqrt{a}} - 1 \right) s_j \right) \mathbb{1}\{m_n = a, \mathcal{A}, \mathcal{B}_a\} \end{aligned} \quad (\text{S128})$$

$$= \left( \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \bar{X}_a^{(n)} + \left( 1 - \frac{\sqrt{a}}{\sqrt{\tilde{m}}} \right) s_j \right) \mathbb{1}\{m_n = a, \mathcal{A}, \mathcal{B}_a\} \quad (\text{S129})$$

We are now ready to prove that (II) = 0.

**Lemma S10** (Nullity of (II)).

$$(II) = 0$$

*Proof.*

$$(II) \tag{S130}$$

$$= \left| \mathbb{E} \left( \sum_{i=1}^{\tilde{m}-1} \left( \Pi_{\hat{j}} \left( \hat{X}_{\hat{j}}^{(n)} \right) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_i^{(n)} - \frac{(\sqrt{\tilde{m}} - \sqrt{i})}{\sqrt{\tilde{m}}} s_{\hat{j}} \right) \mathbf{1}\{\mathcal{A}, \mathcal{B}_i\} \mathbf{1}(m_n = i) \mid \hat{j} \right) \right| \tag{S131}$$

$$= \left| \mathbb{E} \left( \sum_{i=1}^{\tilde{m}-1} \left( \left( \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_i^{(n)} + \left( 1 - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \right) s_{\hat{j}} \right) - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_i^{(n)} - \frac{(\sqrt{\tilde{m}} - \sqrt{i})}{\sqrt{\tilde{m}}} s_{\hat{j}} \right) \mathbf{1}\{A, B_i\} \mathbf{1}(m_n = i) \mid \hat{j} \right) \right| \quad (\text{By eq. (S129)}) \tag{S132}$$

$$= \left| \mathbb{E} \left( \sum_{i=1}^{\tilde{m}-1} \left( \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_i^{(n)} - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \bar{X}_i^{(n)} + \left( 1 - \frac{\sqrt{i}}{\sqrt{\tilde{m}}} \right) s_{\hat{j}} - \left( 1 - \frac{(\sqrt{i})}{\sqrt{\tilde{m}}} \right) s_{\hat{j}} \right) \mathbf{1}\{A, B_i\} \mathbf{1}(m_n = i) \mid \hat{j} \right) \right| \tag{S133}$$

$$= 0 \tag{S134}$$

□

From Lemma S8 and Lemma S10, we obtain

$$\begin{aligned} |\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta| \mathbf{1}\{\mathcal{A}\} &= \frac{\sqrt{\tilde{m}}}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} \frac{1}{\sqrt{\tilde{m}n\alpha^2}} \\ &= \frac{1}{\sqrt{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})n\alpha^2}} \end{aligned} \tag{S135}$$

Incorporating the result from Lemma S5, we get

$$\mathbb{E}[(\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta)^2] \leq \frac{1}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})n\alpha^2} + \frac{4}{\tau} \exp \left( -n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579 \right) \tag{S136}$$

We then deal with the variance term  $\text{Var}(\hat{\theta} \mid \hat{j})$ . Since

$$\hat{\theta} = \frac{\sqrt{\tilde{m}} \frac{2}{n} \sum_{u=n/2+1}^n \hat{\theta}^{(u)}}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} - \frac{\mathbb{E}_{m \sim M}((\sqrt{\tilde{m}} - \sqrt{m}) \mathbf{1}(m \leq \tilde{m}))}{\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})} s_{\hat{j}},$$

and  $(\hat{\theta}^{(u)})_{u=n/2+1}^n$  are independent given  $\hat{j}$ , it holds that

$$\text{Var}(\hat{\theta} \mid \hat{j}) = \frac{2\tilde{m}}{n\mathbb{E}_{m \sim M}(\sqrt{m} \wedge \tilde{m})^2} \text{Var}(\hat{\theta}^{(n)} \mid \hat{j})$$

**Lemma S11** (Majoration of  $\text{Var}(\hat{\theta}^{(n)} \mid \hat{j})$ ).

$$\text{Var}(\hat{\theta}^{(n)} \mid \hat{j}) \leq \frac{2(14\tau)^2}{\alpha^2} \tag{S137}$$

*Proof.*

$$\text{Var}(\hat{\theta}^{(n)} \mid \hat{j}) = \text{Var} \left( \bar{X}^{(n)} + \frac{14\tau}{\alpha} \ell_n \mid \hat{j} \right) \tag{S138}$$

$$= \text{Var} \left( \Pi_{\hat{j}} \left( \bar{X}^{(n)} \right) \right) + \text{Var} \left( \frac{14\tau}{\alpha} \ell_n \right) \tag{S139}$$

$$\leq 14^2 \tau^2 + \frac{14^2 \tau^2}{\alpha^2} \tag{S140}$$

$$\leq \frac{14^2 \cdot 2\tau^2}{\alpha^2} \tag{S141}$$

□



The upper bound of  $\mathbb{E}[(\hat{\theta} - \theta)^2]$  is therefore given by:

$$\mathbb{E}[(\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta)^2] \leq \frac{1}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})n\alpha^2} \quad (\text{S142})$$

$$+ \frac{4}{\tau} \exp\left(-n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579\right) + \frac{\tilde{m}}{(\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}}))^2} \frac{14^2 \cdot 4\tau^2}{n\alpha^2} \quad (\text{S143})$$

$$\leq \frac{(14^2 \times 4 \times 2 + 1) \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}{n\alpha^2 (\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}}))^2} \quad (\text{S144})$$

$$+ 4 \sqrt{\frac{\tilde{m}}{2 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}} \exp\left(-n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579\right) \quad (\text{S145})$$

Recall that  $\tilde{m} = \operatorname{argmax}_{a \in \mathbb{N}^*} \{\mathbb{P}_{m \sim M}(m \geq a) \geq \left(\frac{579}{n\alpha^2} \frac{3}{2} \ln\left(\frac{8(an\alpha^2 \vee 1)}{\ln(8(an\alpha^2 \vee 1))}\right)\right) \wedge 1\}$ . If  $\tilde{m} > 1$ , then it holds that

$$\exp\left(-n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579\right) \leq \frac{\ln(8(\tilde{m}n\alpha^2 \vee 1))^{3/2}}{(8(\tilde{m}n\alpha^2 \vee 1))^{3/2}} \leq \frac{\ln(8(\tilde{m}n\alpha^2 \vee 1))^{3/2}}{8\tilde{m}^{3/2}n\alpha^2},$$

and therefore

$$4 \sqrt{\frac{\tilde{m}}{2 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}} \exp\left(-n\alpha^2 \mathbb{P}(m_u \geq \tilde{m})^2 / 579\right) \leq \frac{\ln(8(\tilde{m}n\alpha^2 \vee 1))}{\tilde{m}n\alpha^2}.$$

As a result, we obtain the bound

$$\begin{aligned} \mathbb{E}[(\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta)^2] &\leq \frac{(14^2 \times 4 \times 2 + 2) \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}{n\alpha^2 (\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}}))^2} \\ &= \frac{1570 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}{n\alpha^2 (\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}}))^2}. \end{aligned} \quad (\text{S146})$$

Otherwise,  $\tilde{m} = 1$  and therefore  $\tau \geq 1$  which has as an immediate consequence that  $\mathbb{P}(\mathcal{A}) = 1$  (there is only one bin so  $\hat{j}$  is necessarily the index of the optimal bin). Then, from the upper bound on  $|\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta| \mathbb{1}\{\mathcal{A}\}$  in Equation (S135), we obtain that

$$\mathbb{E}[(\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta)^2] \leq \frac{1}{\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}})n\alpha^2}. \quad (\text{S147})$$

Incorporating the upper bound on the variance from Lemma S11, we obtain the bound

$$\mathbb{E}[(\mathbb{E}[\hat{\theta} \mid \hat{j}] - \theta)^2] \leq \frac{1569 \ln(8(\sqrt{\tilde{m}n\alpha^2} \vee 1))}{n\alpha^2 (\mathbb{E}_{m \sim M}(\sqrt{m \wedge \tilde{m}}))^2}$$

which together with the bound in Equation (S146) concludes the proof.  $\square$

### S3 Assets used

To conduct our numerical experiments we used Numpy and Statsmodels for computations [Harris et al. \(2020\)](#) [Seabold and Perktold \(2010\)](#) which have a BSD compatible licence and matplotlib for visualisation [Hunter \(2007\)](#) which has a PSF compatible licence.