IEICE TRANS. INF. & SYST., VOL.E105–D, NO.8 AUGUST 2022

PAPER

Locally Differentially Private Minimum Finding

Kazuto FUKUCHI[†,††a)], **Chia-Mu YU**[†††], ***Nonmembers, and*** **Jun SAKUMA**[†,††], ***Member***

SUMMARY We investigate a problem of finding the minimum, in which each user has a real value, and we want to estimate the minimum of these values under the local differential privacy constraint. We reveal that this problem is fundamentally difficult, and we cannot construct a consistent mechanism in the worst case. Instead of considering the worst case, we aim to construct a private mechanism whose error rate is adaptive to the easiness of estimation of the minimum. As a measure of easiness, we introduce a parameter α that characterizes the fatness of the minimum-side tail of the user data distribution. As a result, we reveal that the mechanism can achieve $O((\ln^6 N/\epsilon^2 N)^{1/2\alpha})$ error without knowledge of α and the error rate is near-optimal in the sense that any mechanism incurs $\Omega((1/\epsilon^2 N)^{1/2\alpha})$ error. Furthermore, we demonstrate that our mechanism outperforms a naive mechanism by empirical evaluations on synthetic datasets. Also, we conducted experiments on the MovieLens dataset and a purchase history dataset and demonstrate that our algorithm achieves $\tilde{O}((1/N)^{1/2\alpha})$ error adaptively to α.

key words: *privacy, local differential privacy, minimum finding, estimation error analysis, heavy-tailed distributions*

1. Introduction

Statistical analyses with individuals' data have a significant benefit to our social lives. However, using individuals' data raises a serious concern about privacy, and privacy preservation is increasingly demanding by social communities. For example, the European Commission (EC) approved a new regulation regarding data protection and privacy, the General Data Protection Regulation (GDPR), which has been in effect since May 2018. With this regulation, any service provider in the world must follow GDPR when providing services to any individuals in the EU.

Motivated by the privacy concern, many researchers developed statistical analysis methods with a guarantee of *Differential privacy* [1]. The differential privacy prevents privacy leakage in the central model in which a trusted central server* gathers the individuals' data and then publishes some statistical information about the gathered data to an untrusted analysist. One limitation of this model is that it requires a trusted central server that processes a differentially private algorithm.

A notion of *local differential privacy* (LDP) was introduced by Evfimievski et al. [2] for preventing privacy leakage to the *untrusted* central server. Many researchers proposed some statistical analysis methods with a guarantee of the local differential privacy. For example, mean and median estimation methods [3], distribution estimation [4]–[7], and heavy hitter estimation [8] under the LDP guarantee have been investigated so far.

In this paper, we deal with the *minimum finding problem* under the local differential privacy constraint. Suppose there are N users who have private real-valued data $x_i \in [-1, 1]$ drawn i.i.d. from the distribution whose cumulative distribution is F. An aggregator wants to find the minimum of the users' data $x_{\min} = \inf\{x : F(x) > 0\}$ by collecting the information about the users' private data in the locally differentially private manner. The goal is to construct a locally differentially private estimator of $x_{\min}$ that minimizes the mean absolute error defined as

$$\mathrm{Err} = \mathbf{E}[|\tilde{x} - x_{\min}|], \tag{1}$$

where $\tilde{x}$ denotes the estimated minimum. The minimum finding problem is a primitive but fundamental component for statistical analysis. Even under the privacy constraint, the minimum finding is a necessary first step of statistical analyses.

As we describe later, our mechanism employs binary search to find the interval that contains the minimum. Binary search with local differential privacy has been employed in Gaboardi et al. [9] as a locally differentially private algorithm for estimating the p-quantile of the users' data. They show the minimax optimality of their algorithm in terms of their utility measure, (τ, λ, β)-approximation, up to logarithmic factors, where an estimation algorithm (τ, λ, β)-approximates the p-quantile $x_p = \inf\{x : F(x) > p\}$ if it satisfies either $|\hat{x} - x_p| \le \tau$ or $|F(\hat{x}) - p| \le \lambda$ with probability at least $1 - \beta$. However, upper and lower bounds for the (τ, λ, β)-approximation does not lead bounds on the estimation error in Eq. (1) because we cannot derive any non-trivial bound on Eq. (1) on the event that only the second condition $|F(\hat{x}) - p| \le \lambda$ is satisfied. Further analysis with an additional assumption is necessary to derive the minimax optimal algorithm for minimizing Eq. (1) under the locally differentially private constraint.

Our contributions are listed as follows:

Manuscript received September 15, 2021.
Manuscript revised February 25, 2022.
Manuscript publicized May 11, 2022.

[†]The authors are with the University of Tsukuba, Tsukuba-shi, 305–8577 Japan.
[††]The authors are with Center for Advanced Intelligence Project, RIKEN, Tokyo, 103–0027 Japan.
[†††]The author is with National Yang Ming Chiao Tung University, Taiwan.
a) E-mail: fukuchi@cs.tsukuba.ac.jp

DOI: 10.1587/transinf.2021EDP7187

*The terms *server* and *aggregator* are used interchangeably throughout the paper.

Hardness in the worst case We reveal that the minimum finding problem under the local differential privacy constraint is fundamentally difficult in the worst case. We will prove that no locally differentially private mechanism consistently estimates the minimum under the worst-case users' data distribution.

LDP mechanism with adaptiveness to α-fatness Instead of considering the worst case, we construct a locally differentially private mechanism that is *adaptive to the easiness* of estimation of the minimum, which is determined by the underlying user data distribution. As a measure of easiness, we introduce α-fatness, which characterizes the fatness of the minimum-side tail of the user data distribution. Here, a smaller α indicates that the tail is fatter. The minimum finding problem becomes apparently easier when the underlying distribution is fat because we can expect that a greater portion of data is concentrated around the minimum if the distribution is fatter. Hence, we can expect that the decreasing rate of the estimation error becomes smaller as α decreases. The definition of α-fatness is given as follows:

Definition 1 (α-fatness)**.** *For a positive real α, the distribution of F is α-fat if there exist universal constants $C > 0$ and $\bar{x} \in [-1, 1]$ such that for all $x_{\min} < x < \bar{x}$, $F(x) \geq C(x - x_{\min})^{\alpha}$.*

For example, any truncated distribution, such as the truncated normal distribution, satisfies Definition 1 with $\alpha = 1$. The beta distribution with parameters α and β is α-fat. For simplicity, we say F is α-fat if the F's distribution is α-fat.

Utility analyses We derive adaptive upper bounds on the mean absolute error of the present mechanism as utility analyses and reveal that these bounds are nearly tight. Under the assumption that the server knows a lower bound on α, we show that the mean absolute error is $O((\ln^3 N/\epsilon^2 N)^{1/2\alpha})$, where ϵ is the privacy parameter. If α is unknown to the server, we show that the mean absolute error is $O((\ln^6 N/\epsilon^2 N)^{1/2\alpha})$. Also, we prove that these upper bounds are nearly tight in the sense that any locally differentially private mechanism incurs $\Omega((1/\epsilon^2 N)^{1/2\alpha})$ error under the α-fatness assumption. The error rates of our mechanism become slower as α increases; this reflects the intuition about the easiness of estimation mentioned before. Note that this decreasing rate can be achieved even though the algorithm can use only imperfect knowledge on α (e.g., lower bound on α) or no information about α.

Empirical evaluation We conducted some experiments on real and synthetic datasets for evaluating the performance of the proposed mechanism. In the synthetic datasets experiment, we first confirm the tightness of the theoretical bounds regarding N and ϵ. Furthermore, we demonstrate by the experiment that the present mechanism outperforms a baseline method based on the Laplace mechanism. In the experiment on the real datasets, we evaluate the performance of the proposed mechanism on the MovieLens dataset and a customers' purchase history dataset. As a result, we present that the proposed mechanism succeeds to achieve $\tilde{O}(1/N^{1/2\alpha})$ rate adaptively to α, where the notation $\tilde{O}$ ignores the logarithmic factor.

All the missing proofs can be found in the appendix.

Notations. We denote the indicator function as $\mathbb{1}_x$ for an predicate x. Let $\mathrm{sign}(x) = 1$ if $x \geq 0$, and $\mathrm{sign}(x) = -1$ if $x < 0$. For an event $\mathcal{E}$, we denote its complement as $\mathcal{E}^c$. Let $x_{(1)} \leq x_{(2)} \leq ... \leq x_{(N)}$ be ordered data. We use $\tilde{F}(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{x_i \leq x}$. We define the quantile function of F and $\tilde{F}$ as $F^*(\gamma) = \inf\{\tau : F(\tau) \geq \gamma\}$ and $\tilde{F}^*(\gamma) = \inf\{\tau : \tilde{F}(\tau) \geq \gamma\}$, respectively.

2. Preliminaries

We introduce the fully interactive local differential privacy [10] as a privacy definition; we simply call it the local differential privacy. Suppose that an individual has a data x_i on a domain $\mathcal{X}$. The aggregator conducts an iterative algorithm. For each round t, the aggregator decides whether the algorithm is stopped; if not, they choose a user i_t and then obtains the privatized information Z_t by querying to the i_tth user. Here, the selection of the user i_t can depend on the history $H_t = (i_1, Z_1, ..., i_{t-1}, Z_{t-1})$, and the user calculates the privatized information Z_t from H_t, i_t, and x_{i_t}. The process $Z = (Z_1, Z_2, ..., Z_T)$, where T denotes the round the aggregator says "stop", is dependent on the users' data $X = (x_1, ..., x_N)$; hence, we introduce $Z(X)$ to denote the process of the privatized data when the users' data are X. Then, privacy is defined as follows:

Definition 2 (Local differential privacy [10])**.** *A process of the privatized information $Z(X)$ is ϵ-locally differentially private if for all $X, X' \in \mathcal{X}^N$ differing at most one elemement and all $S \in \sigma(\{Z(X) : X \in \mathcal{X}^N\})$,*

$$\mathbb{P}\{Z(X) \in S\} \leq e^{\epsilon}\mathbb{P}\{Z(X') \in S\}, \tag{2}$$

where $S \in \sigma(\{Z(X) : X \in \mathcal{X}^N\})$ denotes an appropriate σ-field generated from the set of the random variables $Z(X)$.

The parameter ϵ determines a level of privacy; that is, smaller ϵ indicates stronger privacy protection. Roughly speaking, the local differential privacy guarantees that the individual's data cannot be certainly inferred from the privatized data even if an adversary has unbounded computational resources and any prior knowledge.

The mechanism that generates the process Z is more useful if it can carry out under lower interactions between the individuals and aggregator. Hence, we introduce a restriction for the interactions, sequential interactivity [3].

Definition 3 ((Sequentially interactive) local differential privacy [3])**.** *A process $Z(X)$ is ϵ-locally sequential interactive differentially private if $Z(X)$ satisfies Definition 2, and $i_t = t$ for $t = 1, ..., N$ and $T = N$ almost surely.*

In Definition 3, each user i sequentially calculates her privatized information Z_i at once. A user i can utilize the privatized information of the privious users $Z_1, ..., Z_{i-1}$ when she calculates Z_i.

As a simple implementation of the locally differentially private mechanism, the randomized response proposed by Warner [11] is known. This is a mechanism for binary data and outputs a binary value. Let $\mathcal{X} = \mathcal{Z} = \{-1, 1\}$, and let x and z be the individual's data and privatized data by the randomized response, respectively. Then, the randomized response flips the individual's data x with probability $1/1+e^{\epsilon}$, and thus we have $z = x$ with probability $e^{\epsilon}/1+e^{\epsilon}$ and $z = -x$ with probability $1/1+e^{\epsilon}$. This mechanism ensures ϵ-local differential privacy.

Fixed and i.i.d. data settings. While the problem setup described in the introduction employs the *i.i.d.* data setting, we can extend the parts of our results to the different data generation setting, the *fixed* data setting.

(**Fixed data**) The users' data are fixed by some unknown rule.

(**i.i.d. data**) The users' data are drawn i.i.d. from some unknown distribution.

The aggregator in the fixed data setting aims to obtain the minimum among the users' data, whereas they in the i.i.d. data setting aims to obtain the minimum within the support of the underlying users' data distribution.

The unknown rule or distribution is described by a non-decreasing function $F : [0, 1] \to [-1, 1]$. In the fixed data setting, the function F determines the empirical cumulative distribution of the users' data. More precisely, the users' data are determined such that $F(x_{(i)}) = (i-1)/(N-1)$ for all $i = 1, ..., N$. In the i.i.d. data setting, F is the cumulative distribution function of the unknown user data distribution. In the both settings, the minimum of the users' data is defined as $x_{\min} = \inf\{x : F(x) > 0\}$.

3. Algorithm

In this section, we derive an algorithm for the locally private finding minimum problem. To this end, we firstly introduce the non-private version of our minimum finding algorithm (Algorithm 1). Then, we derive our main algorithm (Algorithm 2), which is obtained by privatizing the non-private one.

Algorithm 1 shows the non-private version of the proposed algorithm. It employs the binary search algorithm to find the interval containing the minimum from 2^L distinct intervals obtained by evenly dividing the data domain $[-1, 1]$, where L is some positive integer. More precisely, Algorithm 1 iteratively updates the interval $[\ell_t, r_t]$, where the left-endpoint, midpoint, and right-endpoint of the interval are denoted as ℓ_t, τ_t, and r_t, respectively. In Line 1, Algorithm 1 initializes the first interval $[\ell_1, r_1]$ as the data domain. Then, for each round t, the algorithm halves the interval into $[\ell_t, \tau_t)$ and $[\tau_t, r_t]$ and then chooses either of them that contains the minimum $x_{(1)}$ (in Lines 3-10). After L iterations, the interval becomes the desired one. The algorithm outputs the middle of the interval as the estimated value (in Line 11). Because the length of the last interval is 2^{-L+1} by construction, the error of the estimated value is up to 2^{-L}.

To identify which $[\ell_t, \tau_t)$ and $[\tau_t, r_t]$ contains the minimum, Algorithm 1 first asks each user whether or not his/her data is smaller than τ_t (in Line 4). After that, Algorithm 1 calculates the empirical cumulative distribution at τ_t, $\tilde{F}(\tau_t)$, based on their responses. In Algorithm 1, it is denoted as $\Phi(z)$ in Line 6. Then, $[\ell_t, \tau_t)$ contains the minimum if $\tilde{F}(\tau_t) > 0$, and $[\tau_t, r_t]$ does otherwise.

Algorithm 2 shows the privatized version of Algorithm 1. Algorithm 1 accesses the users' data only through a query that asks whether or not his/her data is smaller than τ_t. We sanitize the query using the randomized response described in Sect. 2 in Line 4 of Algorithm 2. Since the randomized response introduces noise into the query's response, we modify Lines 6 and 7 of Algorithm 1. In Line 6, instead of calculating $\Phi(z)$, Algorithm 2 calculates the unbiased estimated value of $\tilde{F}(\tau_t)$, which is denoted as $\Phi'(z')$. An elementary calculation can confirm the unbiasedness of the estimated value. In Line 7, because $\Phi'(z')$ involves error due to sanitization, we introduce a threshold γ instead of 0.

In Algorithm 2, there are two free parameters; L and γ. We investigate an appropriate choice of L and γ by analyzing the absolute mean error of this algorithm. The results of the analyses are demonstrated in the next section. We remark that due to the binary search strategy in our proposed method, one can easily see that our proposed method can be easily adapted to maximum finding.

4. Analyses

Hardness in the worst case. We first show that the private finding minimum problem is fundamentally difficult. Indeed, we cannot construct a locally differentially private algorithm that consistently estimates the minimum in the worst-case users' data:

Theorem 1. *Suppose ϵ is fixed. In the both setting, for any ϵ-locally differentially private mechanism, there exists F such that* $\mathrm{Err} = \Omega(1)$ *with respect to N.*

From the theorem, we can see that we cannot solve the finding minimum problem with a reasonable utility. In Theorem 1, we consider a situation where the minimum point is isolated to all the other points; that is, $x_1 = -1$ and $x_i = 1$ for $i = 2, ..., N$. The worst-case distribution is not α-fat for any finite α.

Adaptive upper bounds and privacy of Algorithm 2. Next, assuming α-fatness of the user's distribution, we reveal the privacy guarantee and the dependency of the error on ϵ and N regarding Algorithm 2.

Theorem 2. *For any choice of ϵ, L, and γ, Algorithm 2 is ϵ-locally differentially private. Moreover, for some $\alpha > 0$, suppose F is α-fat. For a sequence h_N, let $\gamma = \sqrt{4e^{\epsilon/L}(1+e^{\epsilon/L})h_N/(e^{\epsilon/L}-1)^2 N}$. Then, in both of the fixed and i.i.d. data settings, if $L^2 h_N/\epsilon^2 N = o(1)$, Algorithm 2 incurs an error as*

$$\mathrm{Err} = O\Big(\big(L^2 h_N/\epsilon^2 N\big)^{1/2\alpha} + e^{-h_N} + 2^{-L}\Big). \tag{3}$$

In Theorem 2, there are two free parameters, h_N

Algorithm 1: Non-private finding minimum

Input: Depth L

Initialize $\ell_1 = -1$ and $r_1 = 1$;
for $t = 1$ *to* L **do**
  $\tau_t = \frac{\ell_t + r_t}{2}$;
  Each user reports $z_i = \mathrm{sign}(\tau_t - x_i)$;
  The aggregator obtains $z = (z_1, ..., z_N)$;
  Calculate $\Phi(z) = \frac{1}{2N} \sum_{i=1}^{N} z_i + \frac{1}{2}$;
  if $\Phi(z) > 0$ **then**
    $\ell_{t+1} = \ell_t$ and $r_{t+1} = \tau_t$
  else
    $\ell_{t+1} = \tau_t$ and $r_{t+1} = r_t$
return $\tilde{x} = \frac{\ell_{L+1} + r_{L+1}}{2}$

Algorithm 2: Locally private finding minimum

Input: Depth L and a threshold γ

Initialize $\ell_1 = -1$ and $r_1 = 1$;
for $t = 1$ *to* L **do**
  $\tau_t = \frac{\ell_t + r_t}{2}$;
  Each user reports z'_i obtained by sanitizing $\mathrm{sign}(\tau_t - x_i)$ via randomized response with the privacy parameter ϵ/L ;
  The aggregator obtains $z' = (z'_1, ..., z'_N)$;
  Calculate $\Phi'(z') = \frac{1}{2N} \frac{e^{\epsilon/L}+1}{e^{\epsilon/L}-1} \sum_{i=1}^{N} z'_i + \frac{1}{2}$;
  if $\Phi'(z') \geq \gamma$ **then**
    $\ell_{t+1} = \ell_t$ and $r_{t+1} = \tau_t$
  else
    $\ell_{t+1} = \tau_t$ and $r_{t+1} = r_t$
return $\tilde{x} = \frac{\ell_{L+1} + r_{L+1}}{2}$

and L, which the aggregator should select. We obtain $O((L^2 h_N/\epsilon^2 N)^{1/2\alpha})$ error rate by choosing h_N and L so that the second and third terms in Theorem 2 are lower than the first term.

Let us consider the case where the aggregator has prior knowledge regarding a lower bound on α. In this case, an appropriate choice of h_N and L is shown in the following corollary.

Corollary 1. *For some $\alpha > 0$, suppose F is α-fat. Let $h_N \geq \ln(N)/2\alpha$ and $L = \Theta(\log_2 N)$ such that $L \geq \log_2(N)/2\alpha$. Then, Algorithm 2 incurs an error as*

$$\mathrm{Err} = O\left(\left(\ln^3(N)/\epsilon^2 N\right)^{1/2\alpha}\right). \tag{4}$$

The next corollary is useful if the aggregator does not have any prior information about α. In this case, the decreasing rate of the error is slightly worse than Corollary 1.

Corollary 2. *For some $\alpha > 0$, suppose F is α-fat. Let $h_N = \Theta(\log^2(N))$ and $L = \Theta(\log^2(N))$. Then, Algorithm 2 incurs an error as*

$$\mathrm{Err} = O\left(\left(\ln^6(N)/\epsilon^2 N\right)^{1/2\alpha}\right). \tag{5}$$

As well as the intuition, the decreasing rate becomes faster as α decreases in both settings.

Lower bound for the locally private minimum finding. For confirming tightness of Corollaries 1 and 2, we derive minimax lower bound for the locally private minimum finding.

Theorem 3. *Fix $\epsilon \in [0, 22/35]$, α, and C. In the i.i.d. data setting, for any ϵ-locally sequentially interactive differentially private mechanism, there exists F satisfies Definition 1 with α and C such that for a increasing sequence of N and a decreasing sequence of ϵ,*

$$\mathrm{Err} = \Omega\left((1/\epsilon^2 N)^{1/2\alpha}\right). \tag{6}$$

Remark 1. *Since we prove the privacy of Algorithm 2 by the sequential composition of L times the randomized response, Algorithm 2 satisfies 1-composability introduced by Joseph et al. [10]. Thanks to the result of Joseph et al. [10], we can convert Algorithm 2 to a sequentially interactive ϵ-locally differentially private mechanism without sacrificing the utility. Hence, we can use Theorem 3 as the lower bound for the Algorithm 2.*

As proved in Theorem 3, any locally private mechanism incurs $\Omega((1/\epsilon^2 N)^{1/2\alpha})$ error which matches the upper bounds shown in Corollaries 1 and 2 up to log factors. Note that we derive the lower bound in Theorem 3 in a situation where the aggregator knows the fatness parameter α. If the aggregator does not know α, the minimax error might be greater than the one shown in Theorem 3.

5. Experiment

Here, we present experimental results on synthetic, the MovieLens, and purchase history datasets to show the accuracy advantage of our proposed method and confirm the correctness of our theoretical analysis.

5.1 Synthetic Data

We investigated the error between the real and estimated minimum with synthetic data. The data were generated from a cumulative distribution F according to either of the fixed or i.i.d. data setting. We used the beta distribution to construct F. More precisely, let $[x_{\min}, x_{\min} + \Delta]$ be the support of the data, and let X be a random variable that follows the beta distribution with parameter α and β. Then, F is the cumulative distribution of $x_{\min} + \Delta X$. Δ, a, and b are varied as combination of $\Delta \in \{0.3, 0.6, 0.9\}$, $\alpha \in \{0.5, 0.9, 1, 2, 4\}$, and $\beta \in \{1, 2\}$. For stabilizing an error caused by discritization, we report the worst case mean absolute error among $x_{\min} \in \{0 \times (2-\Delta) - 1, 0.2 \times (2-\Delta) - 1, ..., 1 \times (2-\Delta) - 1\}$. The mean absolute errors were calculated from average of 1000 runs. We also report the 0.05 and 0.95 quantiles of the errors.

We evaluated two different choices of L and h_N corresponding to Corollaries 1 and 2:
(**Lower** α) $L = \lceil \log_2(N)/2 \rceil$ and $h_N = \ln(N)/2$,
(**Unknown** α) $L = \left\lceil \log_2^2(N)/2 \log_2(1000) \right\rceil$ and $h_N =$

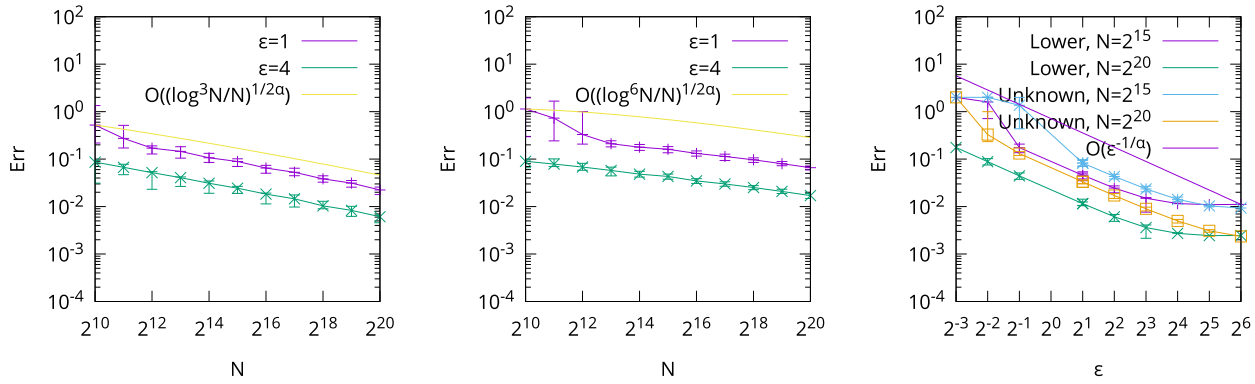


Fig. 1 Err v.s. N (left and middle) and Err v.s. ϵ (right) on the synthetic data. The left figure depicts the result with **Known** α, and the middle figure depicts the result with **Unknown** α.

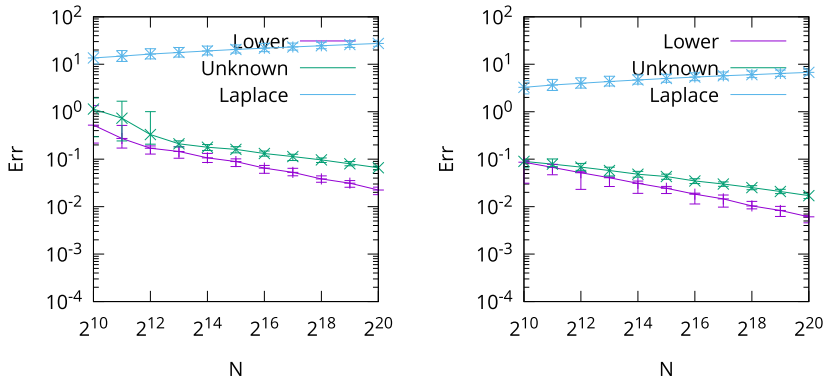


Fig. 2 Comparison between our methods and the baseline method with $\epsilon = 1$ (left) and $\epsilon = 4$ (right).

$\ln^2(N)/2\ln(1000)$.

The lower α case is a suitable parameter choice when the aggregator knows $\alpha \geq 1$, whereas the unknown α case is a suitable parameter choice when the aggregator has no information about α.

Here, we only show partial results in the fixed data setting such that $\alpha = 1, \beta = 1$, and $\Delta = 0.3$. Note that the beta distribution with $\alpha = \beta = 1$ is in fact the uniform distribution.

Error v.s. N) We first demonstrate that Corollaries 1 and 2 are tight with respect to both N. To this end, we evaluated the error of our mechanism corresponding to $N \in \{2^{10}, 2^{11}, ..., 2^{20}\}$.

The left and middle figures in Fig. 1 show the errors of our proposed mechanism with varied N and $\epsilon = 1, 4$. We choose L and h_N according to **Lower** α in the left and **Unknown** α in the middle, respectively. The blue lines denote the theoretical guidelines from Corollaries 1 and 2. We can see from Fig. 1 that the slopes of the errors are almost the same as the slope of the theoretical guideline regardless of choice of ϵ in both **Lower** α and **Unknown** α. This indicates that the decreasing rates with respect to N shown in Corollaries 1 and 2 are tight.

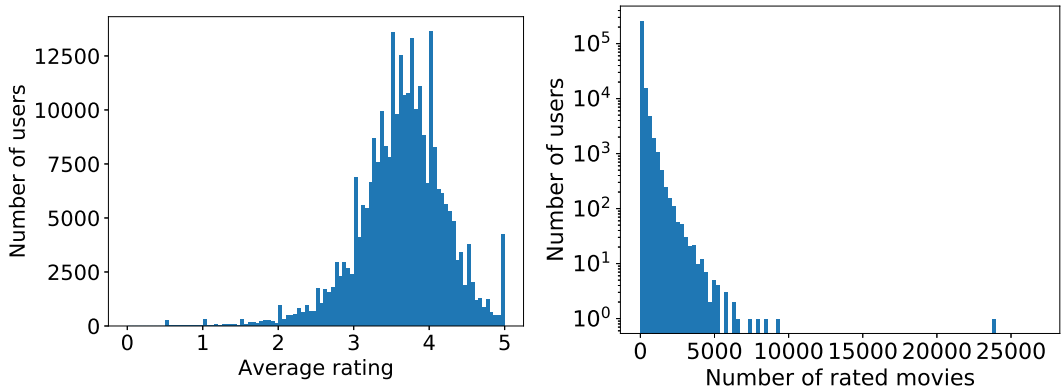
Error v.s. ϵ) Next, we show tightness of Corollaries 1 and 2 regarding ϵ. To this end, we evaluated the error of our mechanism corresponding to $\epsilon \in \{2^{-3}, 2^{-2}, ..., 2^{6}\}$.

The right figure in Fig. 1 shows the errors of our proposed mechanism with varying ϵ and $N = 2^{15}, 2^{20}$. The yellow line represents the theoretical guideline from Corollaries 1 and 2. If ϵ is not large, slopes of the error are almost the same as the slope of the theoretical guideline, where the error is saturated up to 2 for small ϵ since the data are supported on $[-1, 1]$. We therefore can conclude that the rates in Corollaries 1 and 2 with respect to ϵ are tight in the range of small ϵ. Looseness in large ϵ comes from Theorem 2. When deriving Theorem 2, we use a bound $(e^{\epsilon/L}(1+e^{\epsilon/L})/(e^{\epsilon/L}-1)^2)^{1/2\alpha} \leq (2L^2/\epsilon^2)^{1/2\alpha}$, which is valid only if ϵ is sufficiently small. The experimental results reflect this behavior.

In both experiments of error v.s. N and ϵ, the rate looks faster than the theoretical guideline when both N and ϵ are small. This is acceptable because the big-O notation in Theorem 2 indicates that the rate is satisfied only if $L^2 h_N/\epsilon^2 N$ is sufficiently small.

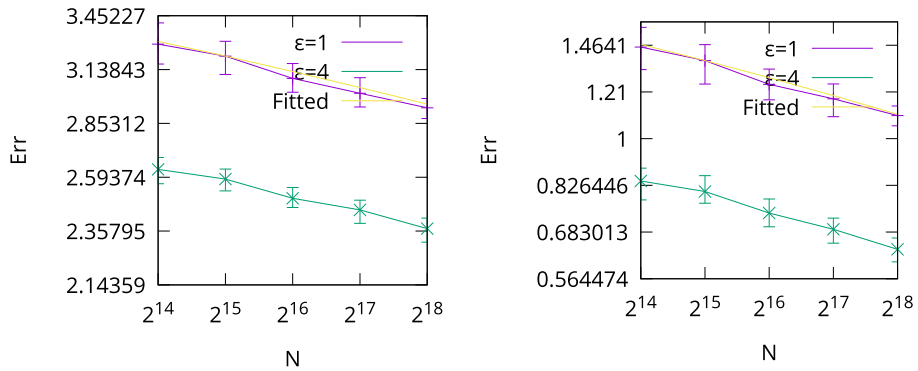
Comparison with Naive Mechanism) We also carried out empirical comparison between our proposed method and a baseline solution. Since there is no existing locally private method for finding the minimum, we consider the straightforward Laplace method as a baseline. In particular, each user with x_i reports $\hat{x}_i = x_i + \delta_i$ with $\delta_i \sim \mathcal{L}(0, 2/\epsilon)$, where $\mathcal{L}(\mu, b)$ is the Laplace distribution with mean μ and scale parameter b. The server simply considers the $\min_i \hat{x}_i$ as the estimated minimum. In this experiment, we use $N \in \{2^{10}, 2^{11}, ..., 2^{20}\}$ and $\epsilon \in \{1, 4\}$.

The comparison between our method and the baseline method is shown in Fig. 2. We can see from Fig. 2 that

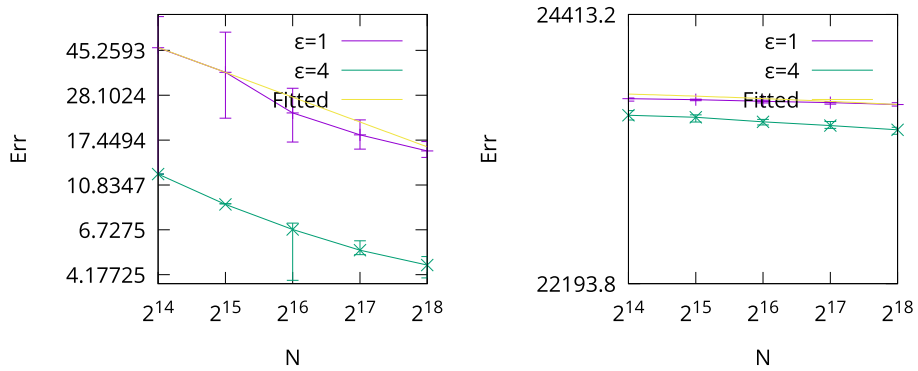


(a) **Task1** (b) **Task2**

Fig. 3 Histogram of the MovieLens dataset for each tasks. Note that the horizontal axis of the right figure is log-scale.



(a) **Task1** min: $\alpha = 7.69$ (b) **Task1** max: $\alpha = 2.50$



(c) **Task2** min: $\alpha = 1.31$ (d) **Task2** max: $\alpha = 113$

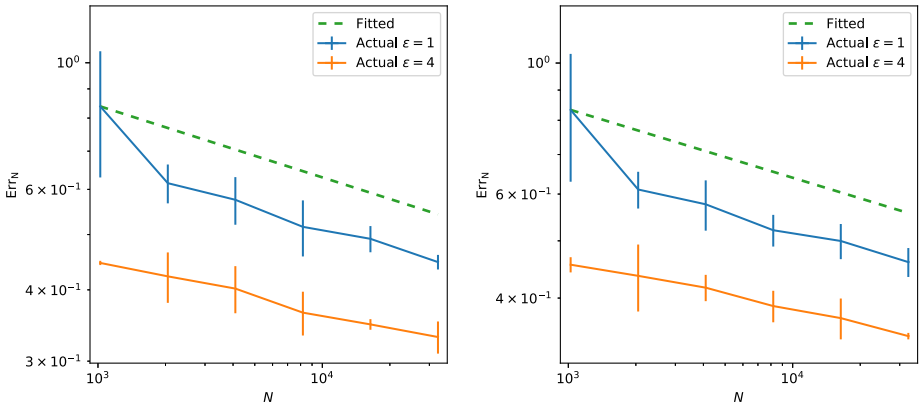
Fig. 4 Err v.s. N on the MovieLens dataset. The yellow line denotes a function $N \to C \log^B N/N^A$ where A and B are obtained by the least square method. We show the value of $\alpha = 1/2A$ in the subcaptions.

the baseline mechanism suffers from an error larger than 1 for all N. Since the data are supported on $[-1, 1]$, the baseline mechanism fails in reasonable estimation. On the other hand, our proposed mechanism achieves significantly smaller error than the baseline method and successes in decreasing its error as N increases.

5.2 MovieLens Data

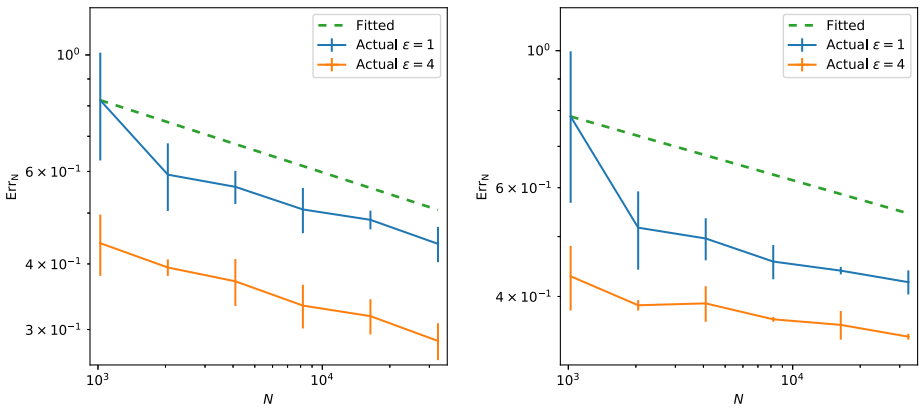
We conducted experiments on the MovieLens dataset[†]. We used the full dataset consisting of 27,753,444 ratings for

[†]Available at https://grouplens.org/datasets/movielens/latest/



(a) **Task1**: $\alpha = 3.98$

(b) **Task1**: $\alpha = 4.31$



(c) **Task2**: $\alpha = 3.60$

(d) **Task2**: $\alpha = 4.79$

Fig. 5 Err v.s. N on the purchase history dataset. The dotted line represents a function $N \to C \log^B N/N^A$ where A and B are obtained by the least square method. We show the concrete value of $\alpha = 1/2A$ in the subcaptions.

53,889 movies obtained by 283,228 users. We carried out the following tasks. (**Taks1**) the server estimates the minimum and maximum of the users' average rating. The domain of the rating is $[0, 5]$. (**Task2**) the server estimates the minimum and maximum numbers of the rated movies per user. We can naturally assume that no user exists that evaluate all the movies. We here assumed that the number of the movies rated by a single user was within [0, 53,889/2]. We evaluated the error of our mechanism with varying $N \in \{2^{14}, ..., 2^{18}\}$ by subsampling the dataset, where we use $\epsilon \in \{1, 4\}$. Since α, the fatness of the distributions, is unknown, we used the **Unknown** α parameter setting shown in Sect. 5.1. The reported value is an average of 1000 runs. We also report the 0.05 and 0.95 quantiles of the errors.

Results) The histograms of the dataset for each task are depicted in Fig. 3. As shown in Fig. 3, the left-side tail of the average review distribution is longer than the right-side tail. Regarding the distribution of the number of reviews per user, the right-side tail is extremely long compared to the left-side tail. We, therefore, can expect that in **Taks 1**, α of the left-side tail is larger than that of the right-side tail, and in **Task 2**, α of the right-side tail is extremely larger than that of the right-side tail.

Figure 4 shows the experimental results. We can see from Fig. 4 that the decreasing rates of the estimation error are changed adaptively to α, and the obtained α shown in the subcaptions corresponds to the fatness of the tail.

5.3 Purchase History Dataset

We also conducted experiments on a purchase history dataset collected in the shopping service provided by Yahoo Japan Corporation. This dataset consists of user attribute information, such as gender and birthday. Also, the dataset contains histories of purchase orders of users in Dec. 2015. Each order consists of a multiset of items purchased. We carried out the following tasks with this dataset:

- **Task1**: The server estimates the minimum age of users whose total amount of purchase on this month was in some range. The ranges are varied as [¥0, ¥10,000], [¥10,000,¥20,000], [¥20,000,¥30,000], [¥30,000,¥40,000], [¥40,000,¥50,000], and [¥50,000,¥60,000].
- **Task2**: The server estimates the minimum age of the users who purchased items in a specific product category.

Here, the age is rescaled from $[0, 150]$ to $[-1, 1]$. The items are categorized into 23 types of products (e.g., fashion, food,

sports), whereas only 19 categories were used so that the number of users who purchased an item in a category is larger than 2^{15}. We evaluated the error of our mechanism with varying $N \in \{2^{11}, ..., 2^{15}\}$ by subsampling the dataset, where we use $\epsilon \in \{1, 4\}$. Since alpha, the fatness of the distributions, is unknown, we used the **Unknown** α parameter setting shown in Sect. 5.1. The reported value is an average of 1000 runs. We also report the 0.05 and 0.95 quantiles of the errors.

Results) Figure 5 shows the experimental results with the real datasets. The figure only consists of the results for **Task 1** with the ranges [¥0, ¥10,000] (left) and [¥40,000, ¥50,000] (right), **Task 2** with the categories music-software (left) and baby-kids-maternity (right).

We can see from Fig. 5 that for these tasks, the estimation error of our proposed mechanism decreases as N increases. Thus, we can expect that our mechanism can consistently estimate the minimum in the real data. Furthermore, the decreasing rates of the estimation error are changed adaptively to the ranges (in **Task 1**) and categories (in **Task 2**).

6. Related Work

LDP gains the first real-world application in Google Chrome's extension, RAPPOR [4] and thereafter also finds applications on the other problems such as distribution estimation [4]–[7] and heavy hitter estimation [8] for categorical-valued data. Different from existing works, our proposed method addresses finding the minimum over numeric-valued data. Simply bucketizing numeric-valued data as categorical data introduces the estimation error. Thus, to handle numeric-valued data, more elaborate protocol design and analysis are needed. There are also local differential privacy methods for numeric-valued problems. For example, Ding et al. [12], Duchi et al. [3], and Nguyen et al. [13] estimate the mean of numeric-valued data under LDP. Ding et al. [14] studied hypothesis testing to compare population means while preserving privacy. Kairouz et al. [15] studied the optimal trade-off between privacy and utility. However, these techniques deal with fundamentally different problems from ours and cannot be extended to the minimum finding problem easily.

Essentially, our proposed method adopts a binary search-based strategy, together with randomized response, to find the minimum. Cyphers et al. [16] developed AnonML to estimate the median over real-valued data under LDP. This method shares the same spirit with ours, i.e., binary search-based strategy with the randomized response. However, the estimation error of their mechanism was not analyzed, for which we cannot set the number of rounds for binary search reasonably.

Our minimum finding mechanism (which can be easily adapted to maximum finding) can be employed as a preprocessing for various types of locally differentially private data analysis. For example, we can use our method for locally differentially private itemset mining [17], [18] over set-valued data. The crucial assumption employed for these methods is that the server knows the maximum number of data items owned by users. Our mechanism can estimate the maximum number in a local differential privacy manner.

7. Conclusion

We investigate the problem of finding the minimum over individuals' data values under local differential privacy. We firstly reveal that this problem is fundamentally hard without any assumption. Hence, we introduce α-fatness for the individuals' data distribution and propose a locally private method for finding minimum under the α-fatness assumption. We reveal that the absolute error of the proposed mechanism is $O((\ln^3 N/\epsilon^2 N)^{1/2\alpha})$ under the α-fatness assumption, as long as the learner knows the value of α. Also, we demonstrate that our mechanism can extend to the case where the learner does not know α, and show that our mechanism achieves $O((\ln^6 N/\epsilon^2 N)^{1/2\alpha})$ adaptively to α. Furthermore, we prove the minimax lower bound of $\Omega((1/\epsilon^2 N)^{1/2\alpha})$, which matches our mechanism's error bounds up to log-factor. The theoretical results imply that a fatter individual's distribution makes the minimum finding problem more difficult. The experimental results demonstrate the tightness of our analyses and applicability of our mechanism to the real-world data.

Acknowledgments

This work was partly supported by KAKENHI (Grants-in-Aid for scientific research) Grant Numbers JP20K19750 and Japan science and technology agency (JST), CREST JPMJCR21D3. Chia-Mu Yu was supported by MOST 111-2636-E-A49-011. We would like to express our gratitude to Yahoo Japan Corporation for providing the purchase history dataset.

References

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," Theory of Cryptography, ed. S. Halevi and T. Rabin, Berlin, Heidelberg, pp.265–284, Springer Berlin Heidelberg, 2006.
[2] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," Proc. twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp.211–222, ACM, June 2003.
[3] J.C. Duchi, M.I. Jordan, and M.J. Wainwright, "Local privacy and statistical minimax rates," 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp.429–438, 2013.
[4] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," Proc. 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS), pp.1054–1067, ACM, Nov. 2013.
[5] G.C. Fanti, V. Pihur, and U. Erlingsson, "Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries," Proceedings on Privacy Enhancing Technologies (PoPETs), vol.2016, no.3, pp.41–61, 2016.
[6] X. Ren, C.M. Yu, W. Yu, S. Yang, X. Yang, J.A. McCann, and P.S. Yu, "LoPub: High-dimensional crowdsourced data publication

with local differential privacy," IEEE Trans. Inf. Forensics Security, vol.13, no.9, pp.2151–2166, Sept. 2018.
[7] T. Murakami and Y. Kawamoto, "Utility-optimized local differential privacy mechanisms for distribution estimation," Proc. 28th USENIX Conference on Security Symposium, SEC'19, USA, p.1877–1894, USENIX Association, 2019.
[8] R. Bassily and A.D. Smith, "Local, Private, efficient protocols for succinct histograms," Proc. Forty-Seventh Annual ACM on Symposium on Theory of Computing (STOC), pp.127–135, ACM, June 2015.
[9] M. Gaboardi, R. Rogers, and O. Sheffet, "Locally private mean estimation: z-test and tight confidence intervals," Proc. Machine Learning Research, ed. K. Chaudhuri and M. Sugiyama, Proc. Machine Learning Research, vol.89, pp.2545–2554, PMLR, 2019.
[10] M. Joseph, J. Mao, S. Neel, and A. Roth, "The role of interactivity in local differential privacy," 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS), pp.94–105, IEEE, Nov. 2019.
[11] S.L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," J. American Statistical Association, 1965.
[12] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems (NIPS), pp.3574–3583, 2017.
[13] T.T. Nguyên, X. Xiao, Y. Yang, S.C. Hui, H. Shin, and J. Shin, "Collecting and analyzing data from smart device users with local differential privacy," ArXiv e-prints, 2016.
[14] B. Ding, H. Nori, P. Li, and A. Joshua, "Comparing population means under local differential privacy: with significance and power," Proc. Thirty-Second AAAI Conference on Artificial Intelligence, pp.26–33, AAAI Press, Feb. 2018.
[15] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," J. Mach. Learn. Res. (JMLR), vol.17, pp.17:1–17:51, 2016.
[16] B. Cyphers and K. Veeramachaneni, "AnonML: Locally private machine learning over a network of peers," IEEE Int. Conf. Data Science and Advanced Analytics (DSAA), pp.549–560, IEEE, 2017.
[17] Z. Qin, Y. Yang, T. Yu, I. Khalil, K. Xiao, and K. Ren, "Heavy hitter estimation over set-valued data with local differential privacy," ACM Conference on Computer and Communications Security (CCS), 2016.
[18] T. Wang, N. Li, and S. Jha, "Locally differentially private frequent itemset mining," IEEE Symposium on Security and Privacy (S&P), pp.127–143, IEEE, 2018.
[19] S. Boucheron, G. Lugosi, and P. Massart, "Concentration inequalities using the entropy method," The Annals of Probability, vol.31, no.3, pp.1583–1614, 7 2003.
[20] J. Duchi, M. Wainwright, and M. Jordan, "Minimax optimal procedures for locally private estimation," J. American Statistical Association, vol.113, no.521, pp.182–201, 2018.

Appendix A: Examples of α-Fat Distributions

For giving a better understanding of α-fatness, we introduce some concrete values of α, C, and $\bar{x}$ for some F.

Example 1 (Beta distribution)**.** *Let X be a random variable following the beta distribution with parameters α and β. Suppose F is the cumulative distribution of $x_{\min} + (x_{\max} - x_{\min})X$. Then, Definition 1 is satisfied with the same α, and with $C = \max\{1, (\alpha\,\mathrm{B}(\alpha,\beta))^{-1}\}/(x_{\max} - x_{\min})^{\alpha}$ and $\bar{x} = x_{\max}$, where $\mathrm{B}(\alpha,\beta)$ denotes the beta function.*

Example 2 (Truncated (normal) distributions)**.** *Suppose F is the cumulative distribution of the truncated normal distribution supported on $[x_{\min}, x_{\max}]$ with parameters $\mu \in [-1, 1]$ and $\sigma^2 > 0$. Then, Definition 1 is satisfied with $\alpha = 1$ and $C = \min\{\phi((x_{\min} - \mu)/\sigma), \phi((x_{\max} - \mu)/\sigma)\}/\sigma(\Phi((x_{\max} - \mu)/\sigma) - \Phi((x_{\min} - \mu)/\sigma)$, where $\phi(x)$ and $\Phi(x)$ denote a density function and cumulative function of the standard normal distribution, respectively. More generally, any truncated distribution satisfies Definition 1 with $\alpha = 1$.*

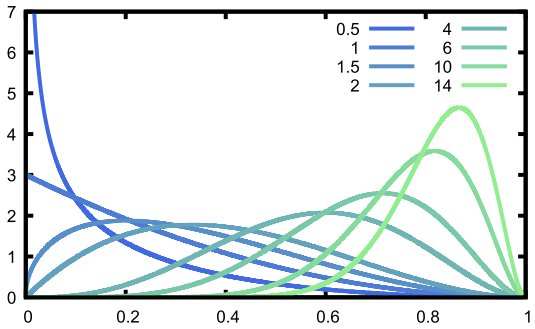


Fig. A· 1 The density function of the beta distribution with $\alpha \in \{0.5, 1, 1.5, 2, 4, 6, 10, 14\}$ and $\beta = 3$.

Figure A· 1 shows the probability density function of the beta distribution with different parameter settings. In Fig. A· 1, we set $\beta = 3$, and α are varied as shown in the legend. We can see from Fig. A· 1 that density around the minimum becomes larger as α decreases.

Appendix B: Analyses of Algorithm 2

B.1 Utility Analysis

Here, we will prove the following two theorems corresponding to the fixed and i.i.d. data settings.

Theorem 4. *Suppose F is α-fat, and γ satisfies*

$$2\gamma < C(\bar{x} - x_{\min})^{\alpha}.$$

In the fixed data setting, for any N, we have

$$\mathrm{Err} \le 2\left(\frac{2\gamma}{C}\right)^{1/\alpha} + \exp\left(-\frac{(e^{\epsilon/L} - 1)^2\gamma^2 N}{4(e^{\epsilon/L} + 1)e^{\epsilon/L}}\right) + 2^{-L}. \quad (\mathrm{A}\cdot 1)$$

Theorem 5. *Suppose F is α-fat, and γ satisfies*

$$2\gamma < C(\bar{x} - x_{\min})^{\alpha}.$$

In the i.i.d. data setting, for any N, we have

$$\mathrm{Err} \le 2\left(\frac{1}{C}\right)^{1/\alpha} \frac{(\lceil 2\gamma N \rceil)^{\overline{1/\alpha}}}{(N+1)^{\overline{1/\alpha}}} + \exp\left(-\frac{(e^{\epsilon/L} - 1)^2\gamma^2 N}{4(e^{\epsilon/L} + 1)e^{\epsilon/L}}\right) + 2^{-L}, \quad (\mathrm{A}\cdot 2)$$

where $(x)^{\overline{\alpha}}$ denotes the rising factorial. That is, letting Γ be the gamma function, $(x)^{\overline{\alpha}} = \Gamma(x + \alpha)/\Gamma(x)$.

With proved Theorems 4 and 5, we can prove Theorem 2.

Proof of Theorem 2. First, we will show that $\gamma = O((L^2 h_N/\epsilon^2 N)^{1/2})$. Because a function $\varphi(x) = e^x(1+e^x)/(e^x-1)^2$ is decreasing and is greather than 1 for $x > 0$, $\varphi(x) = O(1)$ if x is a increasing sequence. If x is a decreasing sequence, we have $\varphi(x) = O(x^{-2})$ because $\varphi(x) \le e(1+e)/x^2$ for $x \in (0, 1)$. Since $\gamma = \sqrt{4\varphi(\epsilon/L)h_N/N}$, we have $\gamma = O((L^2 h_N/\epsilon^2 N)^{1/2})$.

By the assumption, $\gamma = O((L^2 h_N/\epsilon^2 N)^{1/2}) = o(1)$. Hence, γ eventually satisfy the condition $2\gamma < C(\tilde{x} - x_{\min})^\alpha$.

The third terms in Eqs. (A· 1) and (A· 2) match the third term of the bound in Theorem 2. With the choice of γ shown in Theorem 2, we can confirm that the second terms in Eqs. (A· 1) and (A· 2) are $O(e^{-h_N})$, which match the second term of the bound in Theorem 2. Also, since $\gamma = O((L^2 h_N/\epsilon^2 N)^{1/2})$, the first term in Eq. (A· 1) matches the first term of the bound in Theorem 2. Hence, it suffices to prove that the upper bounds on the first term in Eq. (A· 2) matches the first term of the bound in Theorem 2.

We will show that $\frac{(\lceil 2\gamma N\rceil)^{\overline{1/\alpha}}}{(N+1)^{\overline{1/\alpha}}} = O(\gamma)$, by which we can confirm that the upper bounds on the first term in Eq. (A· 2) matches the first term of the bound in Theorem 2. If $\gamma = \omega(1/N)$, we have

$$\lim_{N\to\infty} \frac{(\lceil 2\gamma N\rceil)^{\overline{1/\alpha}}}{(N+1)^{\overline{1/\alpha}}}\left(\frac{N+1}{2\gamma N}\right)^{1/\alpha} = 1.$$

Hence, we have

$$\frac{(\lceil 2\gamma N\rceil)^{\overline{1/\alpha}}}{(N+1)^{\overline{1/\alpha}}} = O\left(\left(\frac{2\gamma N}{N+1}\right)^{1/\alpha}\right) = O(\gamma).$$

□

Here, we give the proof sketch of Theorems 4 and 5. Algorithm 2 can be seen as an algorithm that estimates γ-quantile of the users' data because the algorithm finds $x \in [-1, 1]$ such that $\tilde{F}(x) = \gamma$. Hence, the mean absolute error of Algorithm 2 can be decomposed as

$$\mathrm{Err} \le \mathbf{E}\left[\left|\tilde{F}^*(\gamma) - x_{\min}\right|\right] + \mathbf{E}\left[\left|\tilde{x} - \tilde{F}^*(\gamma)\right|\right]. \tag{A· 3}$$

The first term in Eq. (A· 3) denotes the error between the minimum and γ-quantile, and the second term denotes the estimation error of the γ-quantile.

To analyze the second term in Eq. (A· 3), we define events of *mistake*. For each round t, define an event

$$\mathcal{M}_t = \left\{\tau_t < \tilde{F}^*(\gamma) \implies \Phi(z) \ge \gamma\right\} \cap \left\{\tau_t > \tilde{F}^*(\gamma) \implies \Phi(z) < \gamma\right\}.$$

Then, $\mathcal{M}_t$ represents an event that, at round t, the algorithm chooses an interval that is far from the γ-quantile, and hence we say the algorithm mistakes at round t if $\mathcal{M}_t$ occurs. Then, we obtain the following lemma regarding the estimation error of the γ-quantile:

Lemma 1. *Let τ_t be determined by Algorithm 2. Then, for any random variable $\delta > 0$ that can depend on $x_1, ..., x_N$, we have*

$$\mathbf{E}\left[\left|\tilde{x} - \tilde{F}^*(\gamma)\right|\right] = \delta + \mathbf{E}\left[\max_{t=1,...,L} \mathbb{P}\{\mathcal{M}_t\}\mathbb{1}_{|\tilde{F}^*(\gamma)-\tau_t|>\delta}\right] + 2^{-L}.$$

The concentration inequality gives a bound on the second term in Lemma 1.

Lemma 2. *Let $z = (z_1, ..., z_N)$ be the sanitized version of $(\mathrm{sign}(\tau - x_1), ..., \mathrm{sign}(\tau - x_N))$ using the randomized response with the privacy parameter ϵ. If $\gamma > \tilde{F}(\tau)$,*

$$\mathbb{P}\{\Phi(z) > \gamma\} \le \exp\left(-\frac{(e^\epsilon - 1)^2(\tilde{F}(\tau) - \gamma)^2 N}{4(e^\epsilon + 1)e^\epsilon}\right).$$

Moreover, if $\gamma < \tilde{F}(\tau)$,

$$\mathbb{P}\{\Phi(z) < \gamma\} \le \exp\left(-\frac{(e^\epsilon - 1)^2(\tilde{F}(\tau) - \gamma)^2 N}{4(e^\epsilon + 1)e^\epsilon}\right).$$

Choose δ such that $\tilde{F}^*(2\gamma) - \tilde{F}^*(\gamma) \ge \delta$ or $\tilde{F}^*(\gamma) - \tilde{F}^*(0) \ge \delta$. Then, for any t, $\left|\tilde{F}(\tau_t) - \gamma\right| \ge \gamma$. Thus, we obtain a bound on the second term in Lemma 1 from Lemma 2. We can prove Theorems 4 and 5 by deriving bounds on the first term in Eq. (A· 3) and the first term in Lemma 1, where bounds on these terms can be obtained from Definition 1.

B.2 Privacy Analysis

We can prove the privacy of Algorithm 2 easily with the application of the sequential composition theorem. We confirm that Algorithm 2 ensures ϵ-local differential privacy.

Theorem 6. *Algorithm 2 is ϵ-locally differentially private.*

Proof. Algorithm 2 uses the randomized response L times with privacy parameter ϵ/L. By the sequential composition theorem of the local differential privacy, the total privacy loss is at most ϵ. □

Note that Algorithm 2 is ϵ-locally differentially private for any choice of L and γ.

Appendix C: Proofs

C.1 Proof for Hardness

We introduce the definition of differential privacy for proving Theorem 1. The differential privacy is weaker than the local differential privacy because any analysis satisfying the local differential privacy ensures differential privacy. Thus, if the minimum finding problem is difficult under the differential privacy, the problem is also difficult under the local differential privacy. The formal definition of differential privacy is given as follows:

Definition 4 (Differential privacy [1])**.** *A stochastic mechanism $\mathcal{M}$ mapping from $\mathcal{X}^N$ to $\mathcal{Z}$ is ϵ-differentially private if for all $X, X' \in \mathcal{X}^N$ differing at most one record, and all $S \in \sigma(\mathcal{M}(X))$,*

$$\mathbb{P}\{\mathcal{M}(X) \in S\} \leq e^{\epsilon}\mathbb{P}\{\mathcal{M}(X') \in S\},$$

where $\sigma(\mathcal{Z})$ denotes an appropriate σ-field generated from the random variable $\mathcal{M}(X)$.

Then, we prove Theorem 1.

Proof of Theorem 1. **Fixed data case.** Let F_0 be a cumulative distribution such that $F_0(x) = 0$ for $x \in [-1, 1)$. Let F_1 be another cumulative distribution such that $F_1(x) > 0$ for any $x \in (-1, 1]$. Let X_0 and X_1 be the users' data generated from F_0 and F_1, respectively. Then, X_0 and X_1 have different minimum, whereas the other records are equivalent.

Let $\mathcal{M}$ be a ϵ-differentially private mechanism. Then, its mean absolute errors for X_0 and X_1 are obtained as

$$\begin{aligned}\mathbf{E}[|\mathcal{M}(X_0) - x_{\min}|] =& \mathbf{E}[|\mathcal{M}(X_0) - 1|]\\ \mathbf{E}[|\mathcal{M}(X_1) - x_{\min}|] =& \mathbf{E}[|\mathcal{M}(X_1) + 1|].\end{aligned}$$

Assume

$$\mathbf{E}[|\mathcal{M}(X_0) - 1|] = o(1).$$

Then, by the Markov inequality, we have

$$\mathbf{E}[|\mathcal{M}(X_0) - 1|] \geq \mathbb{P}\{|\mathcal{M}(X_0) - 1| > 1\} = o(1).$$

Because of the differential privacy assumption, we have

$$\begin{aligned}&\mathbb{P}\{|\mathcal{M}(X_0) - 1| > 1\}\\ &\qquad \geq e^{-\epsilon}\mathbb{P}\{|\mathcal{M}(X_1) - 1| > 1\} = o(1).\end{aligned}$$

We obtain a lower bound on the error for X_1 as

$$\begin{aligned}&\mathbf{E}[|\mathcal{M}(X_1) + 1|]\\ \geq&\mathbb{P}\{|\mathcal{M}(X_1) + 1| > 1\}\\ =&1 - \mathbb{P}\{|\mathcal{M}(X_1) - 1| > 1\} = 1 - o(1).\end{aligned}$$

This discussion is true even if we exchange X_0 and X_1. Thus, we obtain the claim.

i.i.d. data case. Let F_0 be the same cumulative distribution above. Let F_1 be a cumulative distribution such that $F_1(x) = \delta$ for any $x \in (-1, 1)$. Note that the distributions of F_0 and F_1 are supported only on $\{-1, 1\}$ such that $\mathbb{P}\{X = -1\} = 0$ under F_0 and $\mathbb{P}\{X = -1\} = \delta$ under F_1. In the similar manner in the fixed data case, assume

$$o(1) = \mathbf{E}[|\mathcal{M}(X_1) + 1|] \geq \mathbb{P}\{|\mathcal{M}(X_1) + 1| > 1\},$$

where the inequality is obtained by the Markov inequality. Since under F_0, all the users' data are 1, the number of the different records between X_0 and X_1 follows the binomial distribution with a parameter N and δ. Let $H(X_0, X_1)$ be the number of the different records between X_0 and X_1. Then, from the differential privacy assumption, we have

$$\begin{aligned}&\mathbb{P}\{|\mathcal{M}(X_1) + 1| > 1\}\\ \geq&\mathbf{E}\Big[e^{-\epsilon H(X_0,X_1)}\mathbb{P}\{|\mathcal{M}(X_0) + 1| > 1|X_0\}\Big]\\ =&\mathbf{E}\Big[e^{-\epsilon H(X_0,X_1)}\Big]\mathbb{P}\{|\mathcal{M}(X_0) + 1| > 1\}\\ \geq&(1 - \delta)^N\mathbb{P}\{|\mathcal{M}(X_0) + 1| > 1\}.\end{aligned}$$

For $\delta = o(1/N)$, we have

$$\begin{aligned}&\mathbb{P}\{|\mathcal{M}(X_1) + 1| > 1\}\\ &\qquad \geq (1 - o(1))\mathbb{P}\{|\mathcal{M}(X_0) + 1| > 1\}.\end{aligned}$$

In the same manner as the fixed data case, we get the claim. □

C.2 Proof for Upper Bounds

We first provide the proofs of Lemmas 1 and 2 and then give the proofs of Theorems 4 and 5.

Proof of Lemma 1. Let $t_1 < t_2 < ... < t_M$ be the rounds that the algorithm mistakes. By the definition of $\mathcal{M}_t$, we have

$$\tilde{F}^*(\gamma) \leq \tau_{t_1} \leq \tau_{t_2} \leq ... \leq \tau_{t_M},$$

or

$$\tilde{F}^*(\gamma) \geq \tau_{t_1} \geq \tau_{t_2} \geq ... \geq \tau_{t_M},$$

Since the algorithm does not mistake after t_M round, we have $|t_M - \tilde{x}| \leq 2^{-L}$. Let t_δ be the maximum round t such that $\left|\tilde{F}^*(\gamma) - \tau_t\right| \leq \delta$. We remark that t_δ is the random variable over $[L]$. Then, we have

$$\begin{aligned}&\mathbf{E}\Big[\left|\tilde{x} - \tilde{F}^*(\gamma)\right|\Big] =\\ &\qquad \mathbf{E}\Big[\left|\tilde{F}^*(\gamma) - \tau_{t_\delta}\right|\Big] + \mathbf{E}\Big[\left|\tau_{t_\delta} - \tau_{t_M}\right|\Big] + 2^{-L}.\end{aligned}$$

Since the difference between τ_t and τ_{t+1} is 2^{-t}, we have

$$\left|\tau_{t_\delta} - \tau_{t_M}\right| \leq \sum_{t=t_\delta+1}^{L} \mathbb{1}_{\mathcal{M}_t} 2^{-t} \leq \sum_{t=1}^{L} \mathbb{1}_{\mathcal{M}_t, \left|\tilde{F}^*(\gamma) - \tau_t\right| > \delta} 2^{-t}.$$

Hence,

$$\begin{aligned}&\mathbf{E}\Big[\left|\tau_{t_m} - \tau_{t_M}\right|\Big]\\ \leq&\mathbf{E}\left[\sum_{t=1}^{L} \mathbb{P}\{\mathcal{M}_t\}\mathbb{1}_{\left|\tilde{F}^*(\gamma) - \tau_t\right| > \delta} 2^{-t}\right]\\ \leq&\mathbf{E}\left[\max_{t=1,...,L} \mathbb{P}\{\mathcal{M}_t\}\mathbb{1}_{\left|\tilde{F}^*(\gamma) - \tau_t\right| > \delta} \sum_{t=1}^{L} 2^{-t}\right]\\ \leq&\mathbf{E}\left[\max_{t=1,...,L} \mathbb{P}\{\mathcal{M}_t\}\mathbb{1}_{\left|\tilde{F}^*(\gamma) - \tau_t\right| > \delta}\right].\end{aligned}$$

□

Proof of Lemma 2. We use the concentration inequality from [19]. Let $Z = f(z_1, ..., z_N) = \frac{e^\epsilon + 1}{e^\epsilon - 1}\sum_{i=1}^{N} z_i$. Let $Z^{(i)} = f(z_1, ..., z_{i-1}, z_i', z_{i+1}..., z_N)$, where z_i' be the independent copy of z_i. Define

$$V_+ = \mathbf{E}\left[\sum_{i=1}^{N}\left(Z - Z^{(i)}\right)^2 \mathbb{1}_{Z > Z^{(i)}} | z_1, ..., z_N\right]$$

$$=\frac{4(e^\epsilon+1)^2}{(e^\epsilon-1)^2}\sum_{i=1}^{N}\mathbb{P}\{z_i'=-1\}\mathbb{1}_{z_i=1}.$$

Moreover, we have

$$\begin{aligned}V_- =&\mathbf{E}\left[\sum_{i=1}^{N}\left(Z-Z^{(i)}\right)^2\mathbb{1}_{Z<Z^{(i)}}|z_1,...,z_N\right]\\ =&\frac{4(e^\epsilon+1)^2}{(e^\epsilon-1)^2}\sum_{i=1}^{N}\mathbb{P}\{z_i'=1\}\mathbb{1}_{z_i=-1}.\end{aligned}$$

From Theorem 2 in [19], for $\theta>0$ and $\lambda\in(0,1/\theta)$, we have

$$\ln\mathbf{E}\left[e^{\lambda(Z-\mathbf{E}[Z])}\right]\le\frac{\lambda\theta}{1-\lambda\theta}\ln\mathbf{E}\left[e^{\frac{\lambda V_+}{\theta}}\right],$$

and

$$\ln\mathbf{E}\left[e^{-\lambda(Z-\mathbf{E}[Z])}\right]\le\frac{\lambda\theta}{1-\lambda\theta}\ln\mathbf{E}\left[e^{\frac{\lambda V_-}{\theta}}\right].$$

By definition, we have

$$\begin{aligned}&\frac{\lambda\theta}{1-\lambda\theta}\ln\mathbf{E}\left[e^{\frac{\lambda V_+}{\theta}}\right]\\ =&\frac{\lambda\theta}{1-\lambda\theta}\sum_{i=1}^{N}\ln\left(\mathbb{P}\{z_i=1\}e^{\frac{4\lambda}{\theta}\frac{(e^\epsilon+1)^2}{(e^\epsilon-1)^2}\mathbb{P}\{z_i=-1\}}\right)\\ =&\frac{4\lambda^2}{1-\lambda\theta}\sum_{i=1}^{N}\frac{(e^\epsilon+1)^2}{(e^\epsilon-1)^2}\mathbb{P}\{z_i=-1\}\\ &+\frac{\lambda\theta}{1-\lambda\theta}\sum_{i=1}^{N}\ln\mathbb{P}\{z_i=1\}.\end{aligned}$$

As $\theta\to0$, we obtain

$$\begin{aligned}&\lim_{\theta\to0}\frac{\lambda\theta}{1-\lambda\theta}\ln\mathbf{E}\left[e^{\frac{\lambda V_+}{\theta}}\right]\\ =&4\lambda^2\frac{(e^\epsilon+1)^2}{(e^\epsilon-1)^2}\sum_{i=1}^{N}\mathbb{P}\{z_i=-1\}\\ \le&4\lambda^2\frac{(e^\epsilon+1)e^\epsilon N}{(e^\epsilon-1)^2}.\end{aligned}$$

Similarly, we obtain

$$\begin{aligned}&\lim_{\theta\to0}\frac{\lambda\theta}{1-\lambda\theta}\ln\mathbf{E}\left[e^{\frac{\lambda V_-}{\theta}}\right]\\ =&4\lambda^2\frac{(e^\epsilon+1)^2}{(e^\epsilon-1)^2}\sum_{i=1}^{N}\mathbb{P}\{z_i=1\}\\ \le&4\lambda^2\frac{(e^\epsilon+1)e^\epsilon N}{(e^\epsilon-1)^2}.\end{aligned}$$

From the Markov inequality, we have

$$\mathbb{P}\{Z>\mathbf{E}[Z]+t\}\le\frac{e^{\lambda(Z-\mathbf{E}[Z])}}{e^{\lambda t}},$$

and

$$\mathbb{P}\{Z<\mathbf{E}[Z]-t\}\le\frac{e^{-\lambda(Z-\mathbf{E}[Z])}}{e^{\lambda t}}.$$

Optimizing λ gives that

$$\mathbb{P}\{Z>\mathbf{E}[Z]+t\}\le\exp\left(-\frac{(e^\epsilon-1)^2t^2}{16(e^\epsilon+1)e^\epsilon N}\right),$$

and

$$\mathbb{P}\{Z<\mathbf{E}[Z]-t\}\le\exp\left(-\frac{(e^\epsilon-1)^2t^2}{16(e^\epsilon+1)e^\epsilon N}\right).$$

Noting that

$$\begin{aligned}&\mathbb{P}\{Z>\mathbf{E}[Z]+t\}\\ =&\mathbb{P}\left\{\frac{e^\epsilon+1}{e^\epsilon-1}\sum_{i=1}^{N}z_i>2N\tilde{F}(\tau)+t\right\}.\end{aligned}$$

Thus, setting $t=2N(\gamma-\tilde{F}(\tau))$ yields the desired claim. □

Proof of Theorem 4. From Eq. (A· 3) and Lemmas 1 and 2, with an appropriate δ, we have

$$\begin{aligned}\mathrm{Err}\le&\mathbf{E}\left[\left|\tilde{F}^*(\gamma)-x_{\min}\right|\right]\\ &+\max\left\{\mathbf{E}\left[\left|\tilde{F}^*(\gamma)-\tilde{F}^*(0)\right|\right],\mathbf{E}\left[\left|\tilde{F}^*(\gamma)-\tilde{F}^*(2\gamma)\right|\right]\right\}\\ &+\exp\left(-\frac{(e^{\epsilon/L}-1)^2\gamma^2N}{4(e^{\epsilon/L}+1)e^{\epsilon/L}}\right)+2^{-L},\end{aligned}\quad(\mathrm{A}\cdot4)$$

where we use the fact that $\tilde{F}^*$ is non-decreasing. The sum of the first two terms is bounded above by

$$2\mathbf{E}\left[\left|\tilde{F}^*(2\gamma)-x_{(1)}\right|\right].\quad(\mathrm{A}\cdot5)$$

If $2\gamma<C_1(C_2-x_{(1)})^\alpha$, we have $\tilde{F}^*(2\gamma)\in(x_{(1)},C_2)$. Hence, under α-fatness, we have $\left|\tilde{F}^*(2\gamma)-x_{(1)}\right|\le\left(\frac{2\gamma}{C_1}\right)^{1/\alpha}$. Substituting this into Eq. (A· 4) yields the desired result. □

Proof of Theorem 5. The proof follows the same manner of that of Theorem 4 except a bound on Eq. (A· 5). Let $U_{(1)},...,U_{(N)}$ be the order statistics of the uniform distribution on $[0,1]$. Then, we have

$$x_{(k)}=F^*(U_{(k)}).$$

Hence,

$$\begin{aligned}&\mathbf{E}\left[\left|\tilde{F}^*(2\gamma)-x_{\min}\right|\right]\\ =&\mathbf{E}\left[\left|F^*(U_{(\lceil2\gamma N\rceil)}-F^*(0)\right|\right]\\ \le&\frac{1}{C_1^{1/\alpha}}\mathbf{E}\left[U_{(\lceil2\gamma N\rceil)}^{1/\alpha}\right].\end{aligned}$$

Since $U_{(k)}$ follows the beta distribution with parameters k and $N-k+1$, we have

$$\begin{aligned}\mathbf{E}\left[U_{(k)}^{1/\alpha}\right]=&\frac{\mathrm{B}(k+\frac{1}{\alpha},N-k+1)}{\mathrm{B}(k,N-k+1)}\\ =&\frac{\Gamma(k+\frac{1}{\alpha})\Gamma(N+1)}{\Gamma(N+1+\frac{1}{\alpha})\Gamma(k)}=\frac{(k)^{\overline{1/\alpha}}}{(N+1)^{\overline{1/\alpha}}}.\end{aligned}$$

□

C.3 Proof for Lower Bound

Proof of Theorem 3. **i.i.d. data case.** We use the lower bound from Duchi et al. [20] for the i.i.d. case.

Theorem 7 ([20]). *Given $\delta > 0$, let F and F' be the cumulative functions such that these minimums, denoted as $x_{\min}$ and $x'_{\min}$, respectively, differs at least 2δ, i.e., $\left|x_{\min} - x'_{\min}\right| \geq 2\delta$. For $\epsilon \in [0, 22/35]$, for any ϵ-locally differentially private mechanism, there exists a cumulative function F_0 such that the error under F_0 is lower bounded as*

$$\mathrm{Err} \geq |\delta|\left(\frac{1}{2} - \sqrt{N\epsilon^2 \mathrm{TV}(F, F')^2}\right),$$

where TV *denotes the total variation distance.*

From Theorem 7, we can obtain a lower bound by designing F and F' so that $\mathrm{TV}(F, F')$ is minimized while satifying $\left|x_{\min} - x'_{\min}\right| \geq 2\delta$ simultaneously. We select different choices of F and F' for $\alpha \in (0, 1)$ and $\alpha \geq 1$.

Case $\alpha \in (0, 1)$. Set

$$F(x) = \begin{cases} (x+1)^\alpha & \text{if } x \in [-1, 0] \\ 1 & \text{otherwise}, \end{cases}$$

$$F'(x) = \begin{cases} 0 & \text{if } x \in [-1, -1+2\delta) \\ (x+1-2\delta)^\alpha & \text{if } x \in [-1+2\delta, 2\delta] \\ 1 & \text{otherwise}. \end{cases}$$

Then, the total variation distance between F and F' is obtained as

$$\begin{aligned} &\mathrm{TV}(F, F') \\ =&1 - \alpha \int_{-1+2\delta}^{0} \min\left\{(x+1)^{\alpha-1}, (x+1-2\delta)^{\alpha-1}\right\} dx \\ =&1 - \alpha \int_{2\delta}^{1} x^{\alpha-1} dx \\ =&1 - (1 - (2\delta)^\alpha) = (2\delta)^\alpha. \end{aligned}$$

Hence, setting $\delta = (16\epsilon^2 N)^{-1/2\alpha}/2$ yields that

$$\mathrm{Err} \geq \frac{1}{8}\left(\frac{1}{16\epsilon^2 N}\right)^{1/2\alpha}.$$

Case $\alpha \geq 1$. Set

$$F(x) = \begin{cases} (x+1)^\alpha & \text{if } x \in [-1, 0] \\ 1 & \text{otherwise}, \end{cases}$$

$$F'(x) = \begin{cases} 0 & \text{if } x \in [-1, -1+2\delta) \\ (x+1)^\alpha - (2\delta)^\alpha & \text{if } x \in [-1+2\delta, 0] \\ 1 & \text{otherwise}. \end{cases}$$

Then, the total variation distance between F and F' is obtained as

$$\mathrm{TV}(F, F') = \int_0^{2\delta} \alpha x^{\alpha-1} dx$$

$$= (2\delta)^\alpha.$$

Hence, with the same setting of δ for $\alpha \in (0, 1)$ case yields the same lower bound. □

Kazuto Fukuchi received a Ph.D. degree from the University of Tsukuba, Tsukuba, Japan, in 2018. He has been an assistant professor in the Faculty of Engineering, Information and Systems, University of Tsukuba, Japan, since 2019. He has also been a visiting researcher at the Center for Advanced Intelligence Project, RIKEN, Japan, since 2019. His research interests include mathematical statistics, machine learning, and their applications.

Chia-Mu Yu is currently an Assistant Professor and Hwa Tse Roger Liang Junior Chair Professor at National Yang Ming Chiao Tung University. He was a visiting scholar at Harvard University, Imperial College London, IBM T. J. Watson Research Center, RIKEN, and TU Darmstadt. Dr. Yu is the recipient of the Young Scholar Fellowship from MOST, K. T. Li Young Researcher Award from ACM/IICM, Observational Research Scholarship from Pan Wen Yuan Foundation, and Project for Excellent Junior Research Investigators from MOST. He was selected as a Junior Program Distinguished Professor at National Chung Hsing University. His research interests include data privacy and information security.

Jun Sakuma received a Ph.D. degree in Engineering from the Tokyo Institute of Technology, Tokyo Japan in 2003. He has been an associate professor in the Department of Computer Science, School of System and Information Engineering, University of Tsukuba, Tsukuba, Japan, since 2009. Prior to that, he worked as an assistant professor in the Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan (2004–2009). He worked as a researcher at Tokyo Research Laboratory, IBM, Tokyo Japan (2003–2004). His research interests include data mining, machine learning, data privacy and secruity.