Context-aware Data Aggregation with Localized Information Privacy

Bo Jiang Ming Li Ravi Tandon
Department of Electrical and Computer Engineering
University of Arizona, Tucson, AZ, USA.
E-mail: {*bjiang, lim, tandonr*}@email.arizona.edu

***Abstract*—In this paper, localized information privacy (LIP) is proposed, as a new privacy definition, which allows statistical aggregation while protecting users' privacy without relying on a trusted third party. The notion of context-awareness is incorporated in LIP by the introduction of priors, which enables the design of privacy-preserving data aggregation with knowledge of priors. We show that LIP relaxes the Localized Differential Privacy (LDP) notion by explicitly modeling the adversary's knowledge. However, it is stricter than 2ϵ-LDP and ϵ-mutual information privacy. The incorporation of local priors allows LIP to achieve higher utility compared to other approaches. We then present an optimization framework for privacy-preserving data aggregation, with the goal of minimizing the expected squared error while satisfying the LIP privacy constraints. Utility-privacy tradeoffs are obtained under several models in closed-form. We then validate our conclusions by numerical analysis using both synthetic and real-world data. Results show that our LIP mechanism provides better utility-privacy tradeoffs than LDP and when the prior is not uniformly distributed, the advantage of LIP is even more significant.**

I. Introduction

Privacy issues are crucial in this big data era, as users' data are collected both intentionally or unintentionally by a large number of private or public organizations. Most of the collected data are used for ensuring high quality of service, but may also put one's sensitive information at potential risk. For instance, when someone is rating a movie, his/her preferences may be leaked; when someone is searching for a parking spot nearby using a smartphone, his/her real location is uploaded and may be prone to leakage. To mitigate such privacy leakage, it is desirable to design privacy-preserving mechanisms that provide strong privacy guarantees without affecting data utility.

Traditional privacy notions such as k-anonymity [1] do not provide rigorous privacy guarantee and are prone to various attacks. On the other hand, Differential Privacy (DP) [2], [3] has become the $defacto$ standard for ensuring data privacy in the database community [4]. The definition of DP assures each user's data has minimal influence on the output of certain types of queries on a database. In the classical DP setting, it is assumed that there is a trusted server which perturbs users' data while answering queries. However, more often then not, organization collecting users' data may not be trustworthy and the database storage system may not be secure [5].

Recently, localized privacy protection mechanisms have gained attention as this setting allows local data aggregation while protecting each user's data without relying on a trusted third party [6]. In localized privacy-preserving data release, each user perturbs his or her data before uploading it; organizations that want to take advantage of users' data then aggregate with collected users' published results. Earliest such mechanism is randomized response [7], which randomly perturbs each user's data. However, the original randomized response does not have formal privacy guarantees. Later, Localized Differential Privacy (LDP) was proposed as a local variant of DP that quantifies the privacy leakage in the local setting. Many schemes were proposed under the notion of LDP. For example, [8]–[10], and Google's RAPPOR [11]. LDP based data aggregation mechanisms have already been deployed in the real-world. For example, in June 2016, Apple announced that it would deploy LDP-based mechanisms to collect user's typing data [12]. However, Tang *et al*. shows that although each user's perturbation mechanism satisfies LDP, the privacy budget is too large ($\epsilon = 43$)[1] to provide any useful privacy protection. Wang *et al*. provide a variety of LDP protocols for frequency estimation [13] and compare their performance with Google's RAPPOR. However, for a given reasonable privacy budget, these protocols provide limited utility. Intuitively, compared with the centralized DP model, it is more challenging to achieve a good utility-privacy tradeoff under the LDP model. The main reasons are two-fold: (1) LDP requires introducing noise at a significantly higher level than what is required in the centralized model. That is, a lower bound of noise magnitude of $\Omega_\epsilon(\sqrt{N})$ is required for LDP, where N is the number of users. In contrast, only $O_\epsilon(1)$ is required for centralized DP [14]. (2) LDP does not assume a neighborhood constraint on users' data as inputs, thus when the domain of data is very large, LDP leads to a significantly reduced utility [15].

In summary, both localized and centralized DP provide strong context-free theoretical guarantees against worst-case adversaries [16]. Context-free means that there is no knowledge of users' data (either instantaneous or statistical). On the other hand, context-aware privacy notions such as ones where statistical knowledge is available are favorable, as the

This work was partly supported by NSF grants CNS-1731164, CNS-1715947, and CCF-1651492.

[1] The parameters, $\epsilon \geq 0$, measures the privacy level. A smaller ϵ corresponds to a higher privacy level.

utility can be increased by explicitly modeling the adversary's knowledge. Information-theoretic privacy notions [17] [18] that incorporate statistical (prior) knowledge fall into this category, which use mutual information (MI) to measure the information leaked about the original database in the released data [19]–[21]. Compared with context-free privacy notions, context-aware privacy notions, especially prior-aware notions achieve a better utility-privacy tradeoff [16].

We next discuss a simple illustrative example to motivate the need and advantages of context-aware privacy notions.

Example 1. *Consider taking a survey over $N = 100$ individuals, where each person is independently asked whether he/she has been infected by some kind of disease. It is known based on clinical studies that this disease infects 1 out of 10 people on average. Each individual holds a local true answer X_i, where i is the individual's index and $X_i = 1$ if his/her answer is yes, $X_i = 0$ if the answer is no. For privacy consideration, each individual perturbs his/her data by a randomized response mechanism (shown in Fig. 1) before publishing it. The goal is to estimate the aggregate $\sum_{i=1}^{N} X_i$ based on $Y_i, i = 1, ..., N$.*

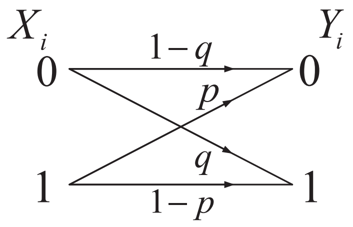


Fig. 1. Randomized response mechanism for each individual.

We first assume that the perturbation mechanism satisfies the context-free ϵ-LDP for a given privacy budget ϵ. By the definition of LDP, for each user: $\max\{\frac{p}{1-q}, \frac{q}{1-p}, \frac{1-q}{p}, \frac{1-p}{q}\} \leq e^{\epsilon}$. In [13], each user's input X_i is treated as a fixed instance and a pair of valid solution for (p, q) is $p = q = \frac{1}{e^{\epsilon}+1}$. Using the unbiased estimator adopted in [13], the expected error of the aggregate is $\mathcal{E}_{LDP} = \frac{Np(1-p)}{(p-q)^2} = \frac{100e^{\epsilon}}{(e^{\epsilon}+1)^2}$. The authors then derived an optimal solution for (p, q) by minimizing this error while subject to the privacy constraints. Their optimal values of p and q are different, where $q^* = 1 - e^{\epsilon}/2$ and $p^* = 0.5$, resulting in an expected estimation error of $\mathcal{E}_{Opt-LDP} = \frac{25}{(0.5-e^{\epsilon}/2)^2}$ in our example, which is smaller than $\mathcal{E}_{LDP}$ (as shown in Fig. 2). This approach increases utility by implicitly using prior knowledge, as it is based on the fact that the answers of a majority of users are zeros (q^* is smaller than p^*). Unfortunately, the assumption that X_i, $i = 1, 2, ..., N$ are instances rather than random variables prohibits introducing prior in a principled manner. In addition, the definition of LDP is independent of the priors, which is unable to adjust the perturbation parameters based on different priors. To explicitly introduce prior knowledge, a new privacy definition is needed.

Assuming that each X_i is a random variable with priors $P_1 = Pr(X_i = 1)$ and $P_0 = Pr(X_i = 0)$, we propose a context-aware LIP notion which imposes a bound on the ratio between the prior and posterior. In this example, we have: $e^{-\epsilon} \leq \{\frac{Pr(Y_i=0)}{1-q}, \frac{Pr(Y_i=0)}{q}, \frac{Pr(Y_i=1)}{1-p}, \frac{Pr(Y_i=1)}{p}\} \leq e^{\epsilon}$.This notion guarantees that taking observations on the published data provides limited additional information of the real data. Restricted by this privacy notion, the optimal p^* and q^* which minimize the mean square error (MSE) can be derived as: $q^* = P_1/e^{\epsilon}$ and $p^* = P_0/e^{\epsilon}$. Intuitively, it reduces MSE by carefully adjusting p and q to different priors. By this perturbation mechanism, the resulting error is $\mathcal{E}_{LIP} = \frac{10e^{\epsilon}-1}{0.64}$ which is significantly smaller than $\mathcal{E}_{LDP}$ (as shown in Fig. 2).

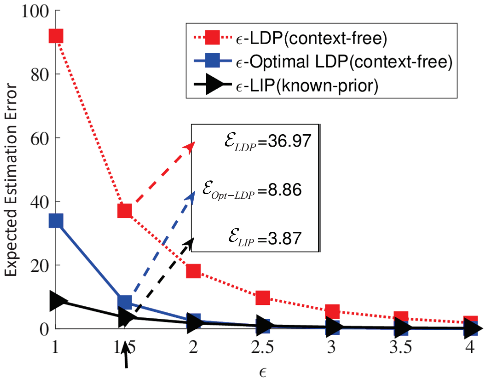


Fig. 2. Comparison among the expected estimation error of different approaches in Example 1: considering prior in the perturbation mechanism significantly reduces the error.

As we discussed above, introducing priors can provide higher utility. However, this comes at the overhead of estimating or learning prior knowledge. This can be obtained in two ways: (1) Sometimes, each individual user's local prior is available (e.g., can be obtained by training based on historical published data) [22], [23]. For instance, when Google wants to survey multiple users' current locations to construct a traffic heat-map, it is possible that it already possesses past reported (unperturbed) locations of each user. (2) However, local-prior is relatively strong knowledge on users' data that may not always be attainable, and one may only be able to learn a global prior (assuming that users' data are identically distributed). For example, when taking a periodic survey, aggregated results in the recent past can be used as the global prior. Another example is when estimating the frequency of a rare disease, one can leverage the results of past clinical research to obtain a global prior [24].

The main contributions of this paper are three-fold:

(1) We propose a new notion of Localized Information Privacy (LIP) for the local data release setting (without a trusted third party), which relaxes the notion of LDP by introducing priors to increase data utility. We formally show that, ϵ-LIP implies ϵ-Mutual Information Privacy (MIP), and ϵ-LIP is sandwiched between ϵ-LDP and 2ϵ-LDP.

(2) We apply the LIP notion to privacy-preserving frequency estimation, and present a utility-privacy optimization framework with the goal of minimizing the mean squared error

while satisfying the LIP constraints. We focus on randomized response type of perturbation mechanisms, and optimal perturbation parameters are derived in closed form under a general binary-input and binary-output model with different local priors, as well as in several special cases. We theoretically demonstrate the advantages of the proposed mechanism.

(3) We validate our analysis by numerical experiments on both synthetic and real-world datasets (i.e., Karosak, a website-click stream data set, and and Gowalla, a location aggregation data set). Both theoretical and numerical results show that optimal perturbation mechanisms under ϵ-LIP always achieve a better utility-privacy tradeoff than those under ϵ-LDP when $\epsilon > 0$, especially when the prior is not uniformly distributed.

The remainder of the paper is organized as follows. In Section II, we describe the proposed LIP metric and its relationship with other existing privacy notions. In Section III, we introduce the system model and problem formulation. In Section IV, we derive the utility-privacy tradeoff under the general model and discuss some special cases. In Section V, we present the numerical results and compare utility-privacy tradeoffs among different models. In Section VI, we offer concluding remarks and discuss future directions.

II. Privacy Definitions

In local privacy-preserving data release, each user uploads its perturbed data directly to an untrusted aggregator. In this section, we first recap two existing privacy notions in localized settings, and then present our new LIP definition.

The traditional LDP definition guarantees that each user's perturbed data has a similar probability to result in the same output for any two inputs from the data domain $\mathbb{D}$:

Definition 1. *(ϵ-Localized Differential Privacy (LDP)) A mechanism $\mathcal{M}$ which takes input X and outputs Y satisfies ϵ-LDP for some $\epsilon \in \mathbb{R}^+$, if $\forall x, x' \in \mathbb{D}$ and $\forall y \in Range(\mathcal{M})$:*

$$\frac{Pr(\mathcal{M}(x) = y)}{Pr(\mathcal{M}(x') = y)} \leq e^{\epsilon}, \tag{1}$$

LDP provides strong context-free privacy guarantee against worst-case adversaries. However, there are many scenarios where some context of X is available (e.g., prior distribution). In such situations, introducing context provides relaxed privacy guarantees. One such definition is mutual information privacy, which uses the mutual information between Y, X to measure the average information leakage of X contained in Y:

Definition 2. *(ϵ-Mutual Information Privacy (MIP)) [20] A mechanism $\mathcal{M}$ which takes input X and outputs Y, satisfies ϵ-MIP for some $\epsilon \in \mathbb{R}^+$, if the mutual information between X and Y satisfies $I(X;Y) \leq \epsilon$, where $I(X;Y)$ is:*

$$\sum_{x,y \in \mathbb{D}} Pr(X = x, Y = y) \log \frac{Pr(X = x, Y = y)}{Pr(X = x)Pr(Y = y)}. \tag{2}$$

Originally, MIP was proposed under the centralized setting where X is the database or individual items and Y is a query output. Here we can adapt it to the local setting, where X and Y are each individual user's input and output. Although MIP is context-aware, it is a relative weak privacy notion since it only bounds the average information leakage. There may exist some (x, y) pair that makes the ratio between the joint and product of marginal distributions very large (while the joint probability is very small). In order to limit the information leakage of every pair of realizations of X and Y, we consider a bound on the ratio between the prior $Pr(X)$ and posterior $Pr(X|Y)$, which leads to our proposed localized information privacy notion:

Definition 3. *(ϵ-Localized Information Privacy (LIP)) A mechanism $\mathcal{M}$ which takes input X and output Y satisfies ϵ-LIP for some $\epsilon \in \mathbb{R}^+$, if $\forall x, y \in \mathbb{D}$:*

$$e^{-\epsilon} \leq \frac{Pr(X = x)}{Pr(X = x|Y = y)} \leq e^{\epsilon}. \tag{3}$$

Intuitively, LIP guarantees that having the knowledge of users' priors, the adversary can't infer too much additional information about each input x by observing each output y. Note that, when ϵ is small, this ratio is bounded close to 1, the mechanism provides strongest privacy guarantee.

Definition. 3 can be viewed as the localized version of information privacy, which focused on a centralized setting [25], and the main differences are in the definitions of input and output. Again, here X and Y stand for each user's input and output variables, respectively.

In the following, We show that LIP is a stronger privacy notion than MIP, since the latter only provides an average privacy guarantee while LIP bounds the leakage on every pair of realizations of X and Y.

Proposition 1. *If a mechanism $\mathcal{M}$ satisfies ϵ-LIP, it also satisfies ϵ-MIP.*

Proof. Assume that $\mathcal{M}$ satisfies ϵ-LIP, by Bayes rules, we have that, $\forall x, y \in \mathbb{D}$:

$$e^{-\epsilon} \leq \frac{Pr(X = x, Y = y)}{Pr(X = x)Pr(Y = y)} \leq e^{\epsilon}. \tag{4}$$

Substituting (4) into (2), we get:

$$I(X, Y) \leq \epsilon \sum_{x,y \in \mathbb{D}} Pr(x, y) = \epsilon,$$

where $\sum_{x,y \in \mathbb{D}} Pr(x, y) = 1$. □

Furthermore, the following theorem shows the relationship between LIP and LDP:

Theorem 1. *If a mechanism $\mathcal{M}$ satisfies ϵ-LIP, then it also satisfies 2ϵ-LDP; if a mechanism $\mathcal{M}$ satisfies ϵ-LDP, then it also satisfies ϵ-LIP.*

Proof. For the first part, consider a mechanism $\mathcal{M}$ that takes any two inputs $X = x$, $X = x'$ and outputs the same $Y = y$.

When $\mathcal{M}$ satisfies ϵ-LIP, using the Bayes rule, Definition 3 is equivalent to:

$$e^{-\epsilon} \leq \frac{Pr(Y = y)}{Pr(Y = y|X = x)} \leq e^{\epsilon}. \tag{5}$$

Since the above also holds for $X = x'$, we have:

$$e^{-\epsilon} \leq \frac{Pr(Y = y)}{Pr(Y = y|X = x')} \leq e^{\epsilon}. \tag{6}$$

Inequality (5) is equivalent to:

$$e^{-\epsilon} \leq \frac{Pr(Y = y|X = x)}{Pr(Y = y)} \leq e^{\epsilon}.$$

Since both of the metrics in above inequalities are positive, by multiplying these two inequalities, we get:

$$e^{-2\epsilon} \leq \frac{Pr(Y = y|X = x)}{Pr(Y = y|X = x')} \leq e^{2\epsilon}.$$

Since we can switch x and x', it is equivalent to the definition of 2ϵ-LDP.

To prove the second part, if $\mathcal{M}$ satisfies ϵ-LDP, we have:

$$Pr(Y = y|X = x') \leq e^{\epsilon} Pr(Y = y|X = x).$$

On the other hand:

$$\begin{aligned} Pr(Y = y) = & \sum_{x' \in \mathbb{D}} Pr(Y = y|X = x')Pr(X = x') \\ \leq & e^{\epsilon} Pr(Y = y|X = x) \sum_{x' \in \mathbb{D}} Pr(X = x') \\ = & e^{\epsilon} Pr(Y = y|X = x), \end{aligned}$$

by switching inputs, we can also get:

$$Pr(Y = y) \geq e^{-\epsilon} Pr(Y = y|X = x),$$

which means that $\mathcal{M}$ also satisfies ϵ-LIP. □

Thus, ϵ-LIP is a more relaxed privacy notion than ϵ-LDP. However, it is stronger than 2ϵ-LDP. Intuitively, LIP relaxes LDP because LDP results in the same output y for every input x, no matter what his/her prior is. On the other hand, for inputs with different priors, LIP perturbs differently. For example, when a user with $Pr(X = 1) = 0.99$, if he holds $X = 1$, which means his real data is consistent with the prior knowledge. As LIP bounds on the ratio between prior and posterior, it has a large probability to output 1 to make the posterior probability similar to its prior; if he holds $X = 0$, which has a small prior to happen, he also has large probability to output 1 to make the posterior probability small.

III. Models and Problem Formulation

A. System and Threat Models

Consider a data aggregation system with N users and a data curator. Each user i locally generates private data which is denoted as random variable X_i, taking value x_i from the domain $\mathbb{D} = \{0, 1, 2...d\}$ with probability $P_k^i = Pr(X_i = k)$. We assume that X_is are independent from each other. Before publishing his/her data to the curator, each user locally perturbs it by a privacy-preserving mechanism $\mathcal{M}_i$. The output is denoted as Y_i which takes value y_i from $\mathbb{D}$. The mechanism $\mathcal{M}_i$ maps each possible input to each possible output with certain probability, the set of them are called perturbation parameters (denoted as $\mathbf{q}^i$). After receiving each user's perturbed data, the curator computes a statistical function on those data (for example, estimating the frequency of certain input which will be useful for data mining). The system model is depicted in Fig. 3.

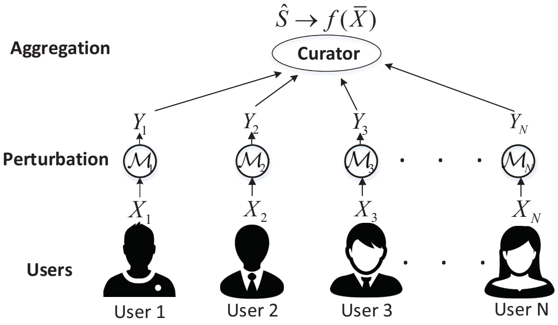


Fig. 3. System Model of Privacy-Preserving Data Aggregation.

The curator is considered as untrusted due to both internal and external threats. On the one hand, users' private data is profitable and companies can be interested in user tracking or selling their data. On the other hand, data breaches may happen from time to time due to hacking activities. Denote the true aggregated result by $f(\bar{X})$, where $\bar{X} = \{X_1, X_2, ..., X_N\}$. The curator (adversary) observes all the users' perturbed outputs $\bar{Y} = \{Y_1, Y_2, ..., Y_N\}$ and tries to obtain an estimate of $f(\bar{X})$. Furthermore, we assume that the adversary possesses knowledge of prior distributions of users' inputs, and the algorithms/perturbation mechanisms that users adopt to publish their data. The curator aims at performing accurate estimations using all the information above, but is also interested in inferring each user's real input X_i. Note that, the curator may have incomplete/less accurate knowledge about each user's local priors. Intuitively, it will reduce the estimation accuracy, however it also mitigates privacy leakage. The utility-privacy tradeoff under such scenarios remain as our future work.

B. Privacy and Utility Definitions

The *privacy* of each user's input satisfies LIP and is parameterized by the privacy budget (ϵ) in Definition 3. The smaller ϵ is, the higher privacy level the mechanism satisfies. Note that, for simplicity, we consider ϵ to be the same for all the users; however it is straightforward to extend our model and results to different ϵ for different users. Under LIP, the privacy constraints can be formulated as:

$$e^{-\epsilon} \leq \frac{Pr(Y_i = y_i|X_i = x_i)}{Pr(Y_i = y_i)} \leq e^{\epsilon}, \forall i \in \{1, 2, ..., N\}, \forall x_i, y_i \in \mathbb{D} \tag{7}$$

Note that, $\mathbf{q}^i \triangleq \{Pr(Y_i = y_i|X_i = x_i), \forall x_i, y_i \in \mathbb{D}\}$. When ϵ is given, the set of inequalities in Eq. (7) constrains $\mathbf{q}^i$ to be within a feasible region $\mathcal{T}_i$, $\forall i \in 1, 2, ..N$.

The definition of *utility* depends on the application scenario. For example, in statistical aggregation, estimation accuracy is often measured by absolute error or mean square error [26]; in location tracking, it is typically measured by Euclidean distance [22]; in privacy-preserving data publishing, distortion is usually used to measure the utility [20]. Since one of the main applications of LIP notion is data aggregation, we define utility as the inverse of the Mean Square Error (MSE): $U(S, \hat{S}) = -\mathcal{E}(S, \hat{S})$, where $\mathcal{E}(S, \hat{S}) = E[(S - \hat{S})^2]$, and $S = f(\bar{X})$, $\hat{S}$ is the estimated S. Thus, maximizing the utility is equivalent to minimizing the MSE.

In general, there is a tradeoff between utility and privacy. We can formulate the following optimization problem to find the optimal perturbation mechanism that yields the optimal tradeoff:

$$\begin{aligned} &\min \mathcal{E}(\mathbf{q}^1, ..., \mathbf{q}^N), \\ s.t. \quad &\mathbf{q}^i \in \mathcal{T}_i, \quad \forall i \in 1, 2, ..N, \end{aligned} \tag{8}$$

where the MSE $\mathcal{E}$ is a function of all the users' perturbation parameters (decision variables), as any estimator $\hat{S}$ will depend on the output $\bar{Y}$ whose distribution is a function of $\mathbf{q}^i$. From [27], it is well known that the optimal estimator that results in the minimized mean square error (MMSE) is $\hat{S} = g(\bar{Y}) = E[S|\bar{Y}]$. Since $E[\hat{S}] = E[S|\bar{Y}] = E[S]$, $\hat{S}$ is an unbiased estimator. Therefore, we use the MMSE estimator in Eq. (8).

C. Frequency Estimation

In this paper, we mainly focus on privacy-preserving frequency estimation. Frequency estimation is one of the most common forms of data aggregation, which is to estimate the percentage of users possessing the value of interest. Example applications include: popularity trend analysis which is used in website click-ratio estimation [13]; identifying heavy-hitters from a stream of sales data or frequently typed keywords on smartphones [28], etc.

Let us assume that the curator wants to survey how many people in the survey possesses a private value v. Each user's data can be encoded into a binary bit as:

$$X_i = \begin{cases} 1 & if \quad v \in user_i \\ 0 & if \quad v \notin user_i. \end{cases} \tag{9}$$

For example, the census bureau wants to know how many people believe in a religion and different users may have different answers. Each user then encodes his/her true data as 1 if he/she believes in that religion, otherwise it is 0.

Currently we only consider binary input data (i.e., $X \in \mathbb{D} = \{0, 1\}$.). If the data domain $\mathbb{D}$ is large, there are many other encoding methods, some of them maps $\mathbb{D}$ into a binary vector, such as unary encoding, direct coding [6], local hashing [29], etc. However, if we consider a vector-type input, we must consider correlations that naturally exist among the bits within the vectors, which will be our future work.

For frequency estimation, the curator needs to estimate the sum $S = \sum_{i=1}^{N} X_i$ using the MMSE estimator $\hat{S} = E[S|\bar{Y}]$. Since we assume that each user's input is independent from each other, $\hat{S}$ can be expressed as:

$$E[\sum_{i=1}^{N} X_i|\bar{Y}] \overset{(a)}{=} \sum_{i=1}^{N} E[X_i|\bar{Y}] \overset{(b)}{=} \sum_{i=1}^{N} E[X_i|Y_i], \tag{10}$$

where (a) in Eq. (10) is due to the independence of X_is, and (b) is because X_i is only correlated with Y_i in the output sequence. Thus $\mathcal{E}(S, \hat{S})$ can be derived as:

$$\mathcal{E}(S, \hat{S}) = E[(\sum_{i=1}^{N} \{X_i - E[X_i|Y_i]\})^2]. \tag{11}$$

Notice that, $\mathcal{E}(S, \hat{S})$ depends on $\mathbf{q}^1, ..., \mathbf{q}^N$ and $P_1^1, ..., P_1^N$, in the context-aware setting, users' priors are assumed to be constants, thus $\mathcal{E}(S, \hat{S}) = \mathcal{E}(\mathbf{q}^1, ..., \mathbf{q}^N)$.

In the local setting, users independently perturb their data, thus each of them results in a MSE in aggregation, which is denoted by $\mathcal{E}_i = E[(X_i - E[X_i|Y_i])^2]$, and the overall utility defined in Eq. (8) satisfies decomposition theorem:

Proposition 2. *For frequency estimation, the global optimization problem defined in Eq.* (8) *can be decomposed into N local optimization problems, under independent user inputs.*

$$\min_{(\mathbf{q}^i) \in \mathcal{T}_i} \mathcal{E}(\mathbf{q}^1, ..., \mathbf{q}^N) = \sum_{i=1}^{N} \min_{(\mathbf{q}^i) \in \mathcal{T}_i} \mathcal{E}_i(\mathbf{q}^i). \tag{12}$$

Proof. The MSE defined in Eq. (11) can be expressed as:

$$\sum_{i=1}^{N} \mathcal{E}_i(\mathbf{q}^i) + 2E[\sum_{k \neq j}^{N} (X_k - E[X_k|Y_k])(X_j - E[X_j|Y_j])]. \tag{13}$$

As users are assumed to be independent, the expected cross terms in Eq. (13) are 0s. Thus $\mathcal{E}(\mathbf{q}^1, ..., \mathbf{q}^N) = \sum_{i=1}^{N} \mathcal{E}_i(\mathbf{q}^i)$.

Assume that for each user, the minimized local MSE $\mathcal{E}_i(\mathbf{q}^i) = e_i$ can be achieved at $\mathbf{q}^{i*}$, where $\mathbf{q}^{i*} \in \mathcal{T}_i$, then $\mathcal{E}(\mathbf{q}^{1*}, ..., \mathbf{q}^{N*}) = \sum_{i=1}^{N} e_i$.

If for some $user_k$ who takes parameters $\mathbf{q}^k \in \mathcal{T}_k$, by assumption, we know that $\mathcal{E}_k(\mathbf{q}^k) \geq e_k$. Thus

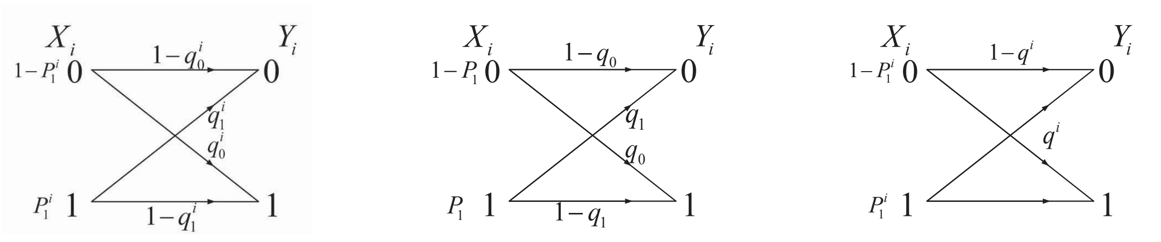
$$\sum_{i=1}^{k} \mathcal{E}_i(\mathbf{q}^{i*}) + \mathcal{E}_k(\mathbf{q}^k) + \sum_{i=k+1}^{N} \mathcal{E}_i(\mathbf{q}^{i*}) \geq \sum_{i=1}^{N} e_i.$$

That means the minimal value of $\mathcal{E}(\mathbf{q}^1, ..., \mathbf{q}^N)$, where $\mathbf{q}^i \in \mathcal{T}_i, \forall i \in [1, N]$ can be achieved if for each user, $\mathbf{q}^i = \mathbf{q}^{i*}$. □

By Proposition 2, when the perturbation parameters of each user are optimal, the overall MSE of the mechanism achieves its minimum. In addition, each user can perform its local optimization independent from each other, which well suits the local setting.

IV. Privacy-Utility Trade-off for Frequency Estimation

In this section, we study the privacy-utility tradeoffs by solving the optimization problems defined in Eq. (8) for frequency estimation. We first analyze a general model where each user has different local priors, and then two special cases are studied (i.e., assuming the same global prior for every user, and a symmetric perturbation model).



(a) The general model: different local priors and asymmetric perturbation

(b) Special case 1: all users have the same prior

(c) Special case 2: symmetric perturbation but different priors

Fig. 4. Different models for the perturbation mechanism considered in this paper (for the i-th user).

A. Utility-Privacy Tradeoff under A General Model

For a general binary input/output model, each user has different local priors and perturbation parameters are asymmetric (shown in Fig. 4(a)). Denote the perturbation parameters by:

$$\begin{aligned} Pr(Y_i = 1|X_i = 0) = q_0^i, \\ Pr(Y_i = 0|X_i = 1) = q_1^i. \end{aligned} \tag{14}$$

Next we derive the concrete optimization objective and constraints. By Eq. (10), the MMSE estimator $\hat{S}_i$ for user i is derived as:

$$\hat{S}_i = E[X_i|Y_i] = P_1^i[\frac{q_1^i}{\lambda_0^i}(1-Y_i) + \frac{1-q_1^i}{\lambda_1^i}Y_i], \tag{15}$$

where $\lambda_0^i = Pr(Y_i = 0) = (1-P_1)(1-q_0^i) + P_1 q_1^i$ and $\lambda_1^i = Pr(Y_i = 1) = (1-P_1)q_0^i + P_1(1-q_1^i)$.

Each user's MSE function $\mathcal{E}_i(q_0^i, q_1^i)$ is derived as:

$$\mathcal{E}_i(q_0^i, q_1^i) = P_1^i(1-P_1^i) - \frac{[P_1^i(\lambda_0^i - q_1^i)]^2}{\lambda_0^i \lambda_1^i}. \tag{16}$$

Detailed derivations are shown in appendix A.

For the privacy constraints, by Eq. (7), when the perturbation mechanism satisfies ϵ-LIP: we have:

$$e^{-\epsilon} \leq \{F_1^i, F_2^i, F_3^i, F_4^i\} \leq e^{\epsilon}, \quad \forall i = 1, 2...N \tag{17}$$

where $F_1^i, F_2^i, F_3^i, F_4^i$ are directly derived from Definition. 3: $F_1^i(q_0^i, q_1^i) = \frac{\lambda_0^i}{q_1^i}$, $F_2^i(q_0^i, q_1^i) = \frac{\lambda_1^i}{1-q_1^i}$, $F_3^i(q_0^i, q_1^i) = \frac{\lambda_0^i}{1-q_0^i}$, $F_4^i(q_0^i, q_1^i) = \frac{\lambda_1^i}{q_0^i}$. Then, the feasible region $\mathcal{T}_i$ is defined as those (q_0^i, q_1^i) pairs satisfying constraints in Eq. (17).

By Proposition 2, the optimization problem of Opt-LIP can be reformulated as:

$$\begin{aligned} \min \mathcal{E}_i(q_0^i, q_1^i), \\ s.t. \ (17), \forall i = 1, 2...N. \end{aligned} \tag{18}$$

We have the following main result:

Theorem 2. *In Opt-LIP, for the i-th user, the optimal* (q_0^i, q_1^i) *pairs that minimize* $\mathcal{E}_i(q_0^i, q_1^i)$ *in problem (18) are: either* $q_0^{i*} = P_1^i/e^{\epsilon}$ *and* $q_1^{i*} = (1-P_1^i)/e^{\epsilon}$, *or* $q_0^{i*} = 1 - P_1^i/e^{\epsilon}$ *and* $q_1^{i*} = 1-(1-P_1^i)/e^{\epsilon}$, *for any given* $\epsilon \geq 0$. *The resulting MSE by* (q_0^{i*}, q_1^{i*}) *is:*

$$\mathcal{E}_{LIP}^* = \sum_{i=1}^{N}\{P_1^i(1-P_1^i)(2e^{-\epsilon} - e^{-2\epsilon})\}. \tag{19}$$

Proof. Here we outline the proof sketch (detailed proofs are shown in Appendix A):

(1) We show that $\mathcal{E}_i(q_0^i, q_1^i)$ is monotonically increasing with q_0^i and q_1^i within the region of $\{q_0^i \geq 0\} \cap \{q_1^i \geq 0\} \cap \{q_0^i + q_1^i \leq 1\}$;

(2) We can simplify the feasible region $\mathcal{T}_i$ by showing that both $\mathcal{E}_i(q_0^i, q_1^i)$ and $\mathcal{T}_i$ are symmetric w.r.t. point $(0.5, 0.5)$; Then we change $\mathcal{T}_i$ to the monotonic region in step. (1).

(3) By the monotonicity, showing that the optimal solution is at the boundary of $\mathcal{T}_i$, which is a linear function of (q_0^i, q_1^i).

(4) The final step is to show that optimal solution is at the intersection of two linear functions in step (3) by testing the monotonicity of $\mathcal{E}_i(q_0^i, q_1^i)$ on the boundary. □

Note that, the optimal solution of each user is achieved when $F_1^i(q_0^i, q_1^i) = F_4^i(q_0^i, q_1^i) = e^{\epsilon}$. Intuitively, to increase utility, we need the probability of perturbation as small as possible (when $q_0^i + q_1^i \leq 0.5$), and the smallest perturbation probability is bounded by the privacy constraints. As a result, the optimal solution is at the point where the privacy requirement is just met. The two optimal (q_0^{i*}, q_1^{i*}) pairs are symmetric w.r.t. $(0.5, 0.5)$. This is due to the symmetric properties of the binary input/output model. The symmetric properties can also be explained as: if we do not consider privacy, utility is maximized in two ways: the first way is each user publishes his/her data directly; the second way is swapping his/her data from 0 to 1 and 1 to 0 before publishing it.

From $q_0^{i*} = P_1^i/e^{\epsilon}$ and $q_1^{i*} = (1-P_1^i)/e^{\epsilon}$, we can see that q_0^{i*} is proportional to P_1^i, and q_1^{i*} is proportional to $1-P_1^i$. Intuitively, from the perspective of one user, when his/her true input value x_i's prior is small, directly revealing x_i will leak too much information about it. In such cases, to satisfy LIP constraints, a large perturbation probability is needed to limit the posterior about x_i. On the contrary, if x_i happens with a large prior, directly releasing x_i will reveal little additional information. Thus a small perturbation probability can be used for x_i in this case.

We next compare our optimal LIP-based perturbation mechanism with the optimal LDP-based one. Define the Opt-LDP problem to be the same with Opt-LIP in (18), except having different privacy constraints of LDP: $\{R_1^i, R_2^i, R_3^i, R_4^i\} \leq e^{\epsilon}$, where $R_1^i, R_2^i, R_3^i, R_4^i$ are derived from Definition 1: $R_1^i = \frac{1-q_1^i}{q_0^i}$, $R_2^i = \frac{q_0^i}{1-q_1^i}$, $R_3^i = \frac{q_1^i}{1-q_0^i}$ and $R_4^i = \frac{1-q_1^i}{q_0^i}$.

Proposition 3. *In Opt-LDP, for the i-th user, the optimal solution for* (q_0^i, q_1^i) *is* $q_0^{i*} = q_1^{i*} = \frac{1}{e^{\epsilon}+1}$, *which results in* *a MSE $\mathcal{E}^*_{LDP}$ of:*

$$\sum_{i=1}^{N}\{P_1^i(1-P_1^i)-\frac{[P_1^i(1-P_1^i)(1-e^\epsilon)]^2}{(1-P_1^i+P_1^ie^\epsilon)(e^\epsilon-P_1^ie^\epsilon+P_1^i)}\}. \quad (20)$$

Given any fixed $\epsilon \geq 0$ and $P_1^i \in [0,1]$, we have $\mathcal{E}^*_{LDP} \geq \mathcal{E}^*_{LIP}$.

Proof. The proof of the optimal solution is similar to that of Opt-LIP, the only difference is the feasible region in Opt-LIP is different for different priors, while the feasible region in Opt-LDP is fixed. The optimal solution also coincides with the one used in [13].

For the second part, it's easy to check by taking derivative over e^ϵ that $\mathcal{E}^*_{LIP} \leq \mathcal{E}^*_{LDP}$, where $\mathcal{E}^*_{LIP} = \mathcal{E}^*_{LDP}$ if $\epsilon = 0$ or $\epsilon = \infty$. This means LIP provides increased utility given any ϵ. We then taking derivative of P_1^i over $\Delta\mathcal{E}^* = \mathcal{E}^*_{LDP} - \mathcal{E}^*_{LIP}$, result shows that $\frac{\partial \Delta\mathcal{E}^*}{\partial P_1^i} = 0$ when $P_1^i = 0.5$. As $\Delta\mathcal{E}^*(P_1^i = 0.5) \geq 0$, $\mathcal{E}^*_{LDP} \geq \mathcal{E}^*_{LIP}$ for any P_1^i. Result also shows that as $|P_1^i - 0.5|$ grows, $\Delta\mathcal{E}^*$ also increases. □

The above result shows that, Opt-LIP always achieves a better utility-privacy tradeoff than Opt-LDP. Intuitively, this is because explicitly considering prior in the privacy definition allows a larger search space than that in Opt-LDP.

B. Utility-Privacy Tradeoff under Special Cases

Next, we investigate two special cases of our general model. Specifically: 1) when each user's local prior are assumed to be the same, which can be substituted by a global prior; 2) each user only perturbs using a symmetric perturbation mechanism, which can be desirable due to less computational complexity.

1) Model of the Same Prior for All Users: When taking a survey of a group of users, it is possible that only the overall distribution of these users' data is obtainable. In this case, we can assume that each user has the same prior (global prior). The perturbation mechanism is shown in Fig. 4(b).

In this model, since q_1^i and q_0^i both depend on P_1^i and the privacy constraints, under a given privacy budget ϵ each user has the same q_1^i and q_0^i. We denote them as q_1 and q_0. We can obtain the following direct result from Theorem 2:

Corollary 1. *In the model in which each user has the same prior, for the i-th user, the optimal (q_0, q_1) pairs that minimize the MSE are: either $q_0^* = P_1/e^\epsilon$ and $q_1^* = (1-P_1)/e^\epsilon$, or $q_0^* = 1 - P_1/e^\epsilon$ and $q_1^* = 1-(1-P_1)/e^\epsilon$, for any given $\epsilon \geq 0$. The resulting minimum MSE is:*

$$\mathcal{E}^*_{GP} = N[P_1(1-P_1)(2e^{-\epsilon} - e^{-2\epsilon})]. \quad (21)$$

We can also see from the result that q_0^* is proportion to P_1 and q_1^* is proportional to $1 - P_1$. The intuition behind this is similar to that of the general model.

2) Model with Symmetric Perturbation Parameters: We also consider a model with binary symmetric input/output perturbation parameters (shown in Fig. 4(c)), where $Pr(Y_i = X_i) = 1 - q^i$ and $Pr(Y_i = 1 - X_i) = q^i$.

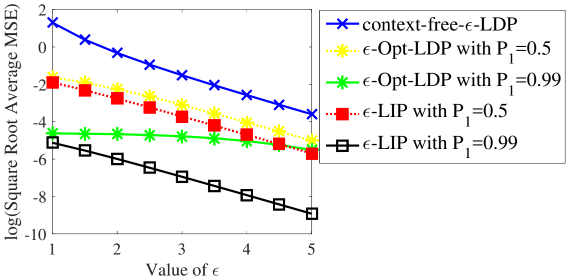


Fig. 5. The utility-privacy tradeoff comparison between prior-aware and prior-free models (log scale for y-axis)

In this case, The MSE $\mathcal{E}_{Sym}$ is derived as:

$$\mathcal{E}_{Sym}(q^i) = \sum_{i=1}^{N}\{P_1^i(1-P_1^i) - {P_1^i}^2\frac{[(2q^i-1)(P_1^i-1)]^2}{(q^i \circledast P_1^i)(1-q^i \circledast P_1^i)}\}, \quad (22)$$

where $q^i \circledast P_1^i = q^i(1-P_1^i) + (1-q^i)P_1^i$.

Proposition 4. *In the model with symmetric perturbation mechanism, for the i-th user, when $P_1^i \leq 0.5$, the optimal q^{i*} that minimizes the MSE is either $q^{i*} = \frac{1-P_1^i}{e^\epsilon - 2P_1^i + 1}$ or $q^{i*} = \frac{e^\epsilon - P_1^i}{e^\epsilon - 2P_1^i + 1}$, for any given $\epsilon \geq 0$; When $P_1^i > 0.5$, the optimal q^{i*} is either $q^{i*} = \frac{e^\epsilon + P_1^i - 1}{e^\epsilon + 2P_1^i - 1}$ or $q^{i*} = \frac{P_1^i}{e^\epsilon + 2P_1^i - 1}$, for any given $\epsilon \geq 0$. The resulting minimum MSE is:*

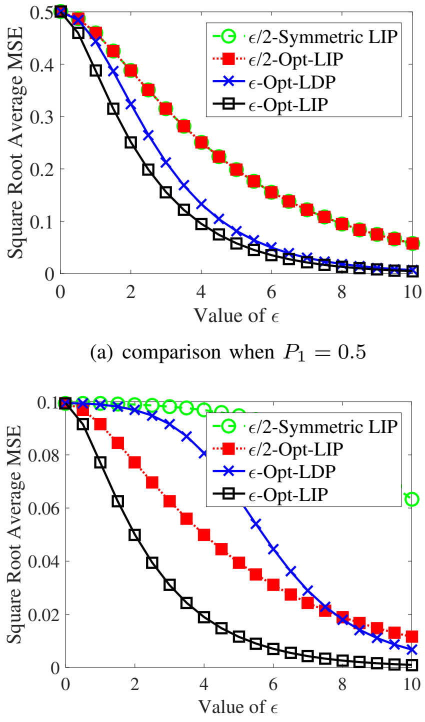
$$\mathcal{E}^*_{Sym} = \sum_{i=1}^{N}\{P_1^i(1-P_1^i) - P_1^i\frac{(1-P_1^i)(2e^\epsilon - e^\epsilon P_1^i - 2P_1^i + 1)^2}{e^{2\epsilon} - 2e^\epsilon P_1^i + e^\epsilon}\}. \quad (23)$$

The proof is a two-dimensional special case of that of Theorem 2. Detailed proof can be found in Appendix B. Note that, when $P_1^i = 0.5$, the optimal perturbation parameters of the symmetric model is same with that of the general model. While this model is simpler to implement, since its parameter search space is smaller, it trades off some utility compared with the general model.

V. Evaluation

In this section, we numerically validate our analytical results. In the first part, we validate our analysis using synthetic data and via Monte-Carlo setup. We first show the advantages of the context-aware privacy notion (based on LIP) versus the context-free notion (based on LDP), by comparing their utility-privacy tradeoffs. Then we compare different models of LIP and LDP under different privacy budgets and prior distributions (with both gloabl priors and local priors). In the second part, we evaluate on two real-world datasets: Karosak (click-streams of websites) and Gowalla (location check-in data).

We evaluate utility by square root average MSE ($\sqrt{\mathcal{E}/N}$). This is because MSE depends on the number of users, for comparison purposes, we first average the MSE to normalize the influence of user count. In addition, since MSE is square error, we take the square root in order to make it comparable



(a) comparison when $P_1 = 0.5$

(b) comparison when $P_1 = 0.01$

Fig. 6. Utility-privacy tradeoffs comparisons with different global prior

with absolute error, which roughly means how much deviation percentage is the result from the exact value. Note that, doing so does not change the optimal solutions in any of our optimization problems. We also use Monte-Carlo simulation to evaluate the convergence behavior of the optimal tradeoff under a large number of users who all have different and random priors. The MSEs are numerical calculated because although simulating the whole randomized response procedure is more meanningful in practice, the numerical results can exactly measure the expectation. In addition, the MSE calculated by simulation converges to the numerical results when the number of simulation instances is large. Since LIP achieves a relaxed privacy guarantee than LDP, it is difficult to compare their utilities under the same privacy level. Thus, we will compare their optimal utilities under any given privacy budget ϵ. All the analysis are done in Matlab (R2016a) on a Dell desktop (OptiPlex 7040; CPU: Intel (R) Core (TM) i5-6500 @ 3.2GHz; RAM 8.0 GB; OS: windows 64bit).

A. Numerical Analysis on Synthetic Data

1) Benefit of Context-awareness: First we would like to compare the utility-privacy tradeoffs between ϵ-LIP and ϵ-LDP, and the goal is to observe the advantage of our proposed context-aware notion versus context-free notion of LDP. Intuitively, the utility gain of the former can be attributed to two factors: 1) using the prior in the MMSE estimator, which improves the accuracy compared with estimators that do not use prior knowledge; 2) the privacy guarantee of LIP is relaxed compared with LDP, by explicitly modeling prior in the definition. As a result, less perturbation is needed to satisfy the same privacy budget ϵ. The latter factor is already proven in Proposition 3. To decouple the influence of the above two factors, we compare the utility-privacy tradeoff of Opt-LIP with two other schemes: Opt-LDP (defined earlier), as well as context-free LDP adopted in previous work [13]. The prior-unaware estimator used in [13] is denoted as $\hat{C}$, which treats X_i as instances rather than variables:

$$\hat{C} = \frac{\sum_{i=1}^{N} Y_i - Np_i}{1 - 2p_i}, \tag{24}$$

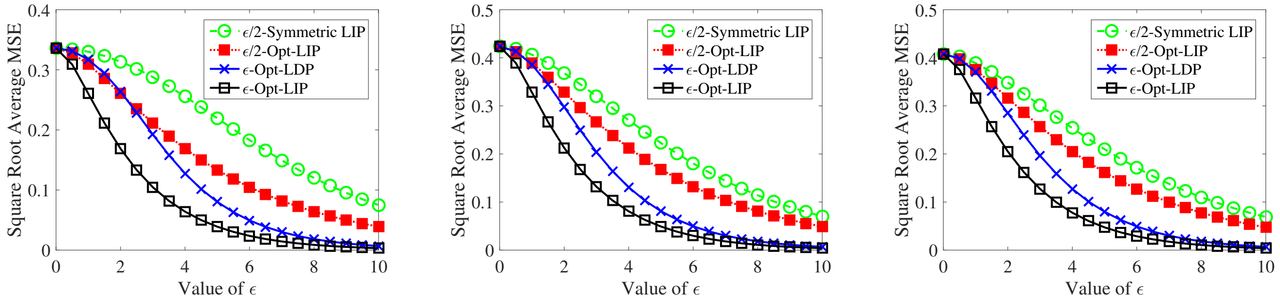
where $p_i = \frac{e^\epsilon}{e^\epsilon + 1}$ as we discussed in section IV. This is an unbiased estimator under the binary symmetric channel (BSC) model. The MSE function of this estimator is:

$$E[(S - \hat{C})^2] = Var[\hat{C}] = \frac{Np_i(1 - p_i)}{(1 - 2p_i)^2} \tag{25}$$

The comparison is shown in Fig. 5. The privacy budget ϵ changes from 1 to 5 with a step of 0.5. For now we assume that N users share the same global prior. We can see that, the square root average MSE of "ϵ-Opt-LDP" is always smaller than that of "context-free LDP" under any given ϵ. When $P_1 = 0.5$ (prior is uniformly distributed), the distance between these two models is smaller; when the prior is more skewed, advantage of the former is even enhanced. This validates the benefit of the prior-aware estimator. On the other hand, by comparing the curves of "ϵ-Opt-LDP" and "ϵ-Opt-LIP" (using the same MMSE estimator), the error of Opt-LIP is always smaller than that of Opt-LDP, and the gap between the two models increases when $P_1 = 0.99$. This result confirms that our relaxed prior-aware privacy notion leads to increased utility. Since the context-free LDP provides the worst utility, we only focus on prior-aware notions and estimators in the following.

2) Comparing Different Prior-Aware Models: To evaluate the utility-privacy tradeoffs of different prior-aware models under different priors and privacy levels, we generate synthetic data in which N users share a global prior. We change ϵ from 0 to 10 with a step of 0.5 and fix the global prior as $P_1 = 0.5$ and $P_1 = 0.01$ respectively. Fig. 6 shows the comparison among four different models: ϵ-Opt-LDP (with the MMSE estimator), ϵ-Opt-LIP, $\epsilon/2$-Opt-LIP and $\epsilon/2$-symmetric LIP. From the curves of "$\epsilon/2$-symmetric LIP" and "$\epsilon/2$-Opt-LIP", we can see that when $P_1 = 0.5$, these two models are equivalent; when $P_1 = 0.01$, the general model performs better than symmetric model. It shows that the ϵ-Opt-LIP model results in smaller errors than the ϵ-Opt-LDP model under any priors and ϵ. When $P_1 = 0.01$, the ϵ-Opt-LIP model performs even better. The result also shows that the curve of "ϵ-LDP" is almost bounded within the region between $\epsilon/2$-Opt-LIP and ϵ-Opt-LIP for $P_1 = 0.5$. But when $P_1 = 0.01$, for smaller values of ϵ ($\epsilon \leq 8.3$), $\epsilon/2$-Opt-LIP performs even better than ϵ-LDP. This means for smaller ϵ values, $\epsilon/2$-Opt-LIP provides both better utility and stronger privacy guarantee than ϵ-Opt-LDP.

One remarkable insight is, even when $P_1 = 0.5$ (the prior knowledge provides the least help to data aggregation), ϵ-



(a) N=5

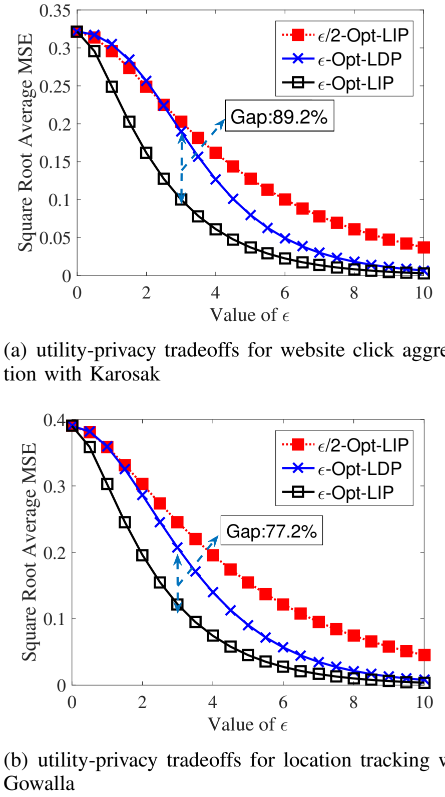
(b) N=500

(c) N=50000

Fig. 7. Utility-privacy tradeoff comparison when N users have different local priors, y-axis is square root average MSE: $\sqrt{\mathcal{E}/N}$.

Opt-LIP still outperforms ϵ-Opt-LDP. This is because for any ϵ, LIP guarantees a weaker privacy level than LDP (i.e., LIP still provides larger feasible regions for the perturbation parameters). When $P_1 = 0.01$, users' inputs are highly certain, merely considering prior in the estimator can already result in accurate aggregation. Thus even $\epsilon/2$-LIP can provide better utility than ϵ-LDP.

3) Monte-Carlo Simulation: We further study the case when each user has different local priors and use Monte-Carlo Simulation to study the convergence of performance when N increases. Fig. 7 shows the comparison among four models described above, where each user's prior probability is sampled uniformly at random from $[0, 1]$. We set $N = 5$, 500 and 50000 and run each simulation once (there is no need to average on multiple simulation runs of each instance, as each user is assumed independent from each other and generating multiple datasets is equivalent to enlarge N). Note that when $N = 5$, curves of $\epsilon/2$-Opt-LIP and ϵ-Opt-LDP cross over. This is because some users' P_1^i values are far from 0.5, which makes the MSE of $\epsilon/2$-Opt-LIP smaller than that of ϵ-Opt-LDP for smaller ϵ. As N becomes larger, we can see that ϵ-Opt-LIP always provides higher utility than ϵ-Opt-LDP model under any ϵ. We also observe that the curve of ϵ-Opt-LDP is almost bounded between the ones of $\epsilon/2$-Opt-LIP and ϵ-Opt-LIP.



(a) utility-privacy tradeoffs for website click aggregation with Karosak

(b) utility-privacy tradeoffs for location tracking with Gowalla

Fig. 8. Utility-privacy tradeoff comparisons using real world data

B. Numerical Analysis with Real-world Datasets

1) Website Popularity Statistics: The first comparison is run using the dataset of Karosak which is a collection of anonymized click-stream data of a hungarian on-line news portal [13]. There are around 8 million click events for 41,270 different pages. In this data set, each row stands for a click-stream for a website in different time slots. Our goal is to estimate the frequency of popular websites (with a total click over 15,000). We treat each website as a user, thus $X_i = 1$ if the total clicks of website i is above 15,000; otherwise, $X_i = 0$. Since no historical data is available, we regard it as the first special case in which users have a global prior.

In Fig. 8(a), we can see that ϵ-Opt-LDP results in larger square root average MSE than ϵ-Opt-LIP. For some smaller value of ϵ, $\epsilon/2$-Opt-LIP performs better than ϵ-Opt-LDP. This is because the popular websites are rare, thus the global prior is relative small. Again, this result confirms that, the more specific the prior is, the more beneficial is LIP than LDP.

2) Location Check-In Dataset: We then compare the performance of different models with another real-world dataset Gowalla, which is a social networking application where users share their locations by checking-in. There are 6,442,892 users in this dataset. For each user, a trace of his/her check-in locations are recorded. For this dataset, we wish to estimate the frequency of users' last check-in location in a specific region v. We first divide the area into 36×36 districts and select one of them as the "secret" region v with coordination from (50,0) to (55,-5) (the area of v covers most of the United Kingdom). We then map each user's locations into districts, and encode them as 1 if a location is in v, and 0 if not.

We obtain each user's local prior by training: we calculate the percentage of times that he/she has been to this location before the last reported location (we ignore user-user correlations). Then we use the last reported location as the real input data. Results are shown in Fig. 8(b), where similar trends can be observed as in the previous datasets.

VI. Conclusion and Future Work

In this paper, the notion of localized information privacy is proposed. As a context-aware privacy notion, it provides relaxed privacy guarantee than LDP by introducing prior knowledge in the privacy definition while achieving increased utility. Combined with an MMSE estimator which also leverages prior knowledge, larger gains in utility can be obtained. We studied the utility-privacy tradeoff of our proposed LIP notion and the traditional LDP notion. We show that our ϵ-LIP always outperform ϵ-LDP for any given privacy budget ϵ. The advantage is more enhanced when the prior distribution is more skewed (even $\epsilon/2$-LIP with a stronger privacy guarantee than ϵ-LDP is better than the latter for small ϵ values).

For future work, we will extend our work to handle more general models. For example, we will consider the multiple-input and multiple-output perturbation model for a large domain, consider an adversary that has less accurate/complete prior knowledge than users, and also understand the impact of user correlations. We will also study the optimality of the binary-input and binary-output perturbation model itself, by comparing it with the binary-input and multiply-output model.

Appendix A
Proof of theorem 4

Proof.

1) Step 1

In order to derive the MSE function that the optimal estimator results in, by the law of total variance:

$$\begin{aligned}\mathcal{E}_i(q_0^i, q_1^i) = E[(X_i - \hat{S}_i)^2] =& E[Var(X_i|\bar{Y})]\\ =& Var(X_i) - Var[E(X_i|\bar{Y})]\\ =& Var(X_i) - Var[\hat{S}_i].\end{aligned} \tag{26}$$

by (26), the remaining problem is to find the value of $Var(X_i)$ and $Var(\hat{S}_i)$. In the prior-aware setting, $Var(X_i) = P_1^i(1 - P_1^i)$. On the other hand,

$$\begin{aligned}Var(\hat{S}_i) =& Var\{P_1^i[\frac{q_1^i}{\lambda_0^i}(1 - Y_i) + \frac{1 - q_1^i}{\lambda_1^i} Y_i]\}\\ =& \frac{[P_1^i(\lambda_0^i - q_1^i)]^2}{\lambda_0^i \lambda_1^i}.\end{aligned} \tag{27}$$

Take (27) into (26), the each user's MSE function $\mathcal{E}_i(q_0^i, q_1^i)$ can be derived

$$\mathcal{E}_i(q_0^i, q_1^i) = P_1^i(1 - P_1^i) - \frac{[P_1^i(\lambda_0^i - q_1^i)]^2}{\lambda_0^i \lambda_1^i}. \tag{28}$$

Notice that P_1^i is a constant, thus the optimization problem is equivalent to maximize:

$$L^i(q_0^i, q_1^i) = \frac{(\lambda_0^i - q_1^i)^2}{\lambda_0^i \lambda_1^i} = \frac{(\lambda_1^i q_1^i - \lambda_0^i(1 - q_1^i))^2}{\lambda_0^i \lambda_1^i}. \tag{29}$$

In order to test the monotoncity of $L^i(q_0^i, q_1^i)$, taking partial derivative on it.

$$\begin{aligned}&\frac{\partial L^i(q_0^i, q_1^i)}{\partial q_0^i}\\ =&(1 - q_1^i)^2(\frac{\lambda_0^i}{\lambda_1^i})' + {q_1^i}^2(\frac{\lambda_1^i}{\lambda_0^i})' + [-2q_1^i(1 - q_1^i)]'\\ =&(1 - q_1^i)^2(\frac{P_1^i - 1}{{\lambda_1^i}^2}) + {q_1^i}^2(\frac{1 - P_1^i}{{\lambda_0^i}^2})\\ =&(1 - P_1^i)(\frac{(\lambda_1^i q_1^i)^2 - [\lambda_0^i(1 - q_1^i)]^2}{{\lambda_1^i}^2 {\lambda_0^i}^2})\\ =&(1 - P_1^i)\frac{[\lambda_1^i q_1^i + \lambda_0^i(1 - q_1^i)][\lambda_1^i q_1^i - \lambda_0^i(1 - q_1^i)]}{{\lambda_1^i}^2 {\lambda_0^i}^2}\\ =&(1 - P_1^i)^2\frac{[\lambda_1^i q_1^i + \lambda_0^i(1 - q_1^i)](q_1^i + q_0^i - 1)}{{\lambda_1^i}^2 {\lambda_0^i}^2}.\end{aligned} \tag{30}$$

In (31), noticed that $\frac{\partial L_i(q_0^i, q_1^i)}{\partial q_0^i} \geq 0$ when $q_1^i + q_0^i - 1 \geq 0$; $\frac{\partial L_i(q_0^i, q_1^i)}{\partial q_0^i} < 0$ when $q_1^i + q_0^i - 1 < 0$. Which means given any value of q_1^i, $L^i(q_0^i, q_1^i)$ is monotonically decreasing with q_0^i from 0 to $1 - q_1^i$.

By the same way, when q_0^i is fixed, $L_i(q_0^i, q_1^i)$ is monotonically decreasing with q_1^i from 0 to $1 - q_0^i$.

2) Step 2

To test the symmetry of $L_i(q_0^i, q_1^i)$, take $q_0^{i'} = 1 - q_0^i$, $q_1^{i'} = 1 - q_1^i$.

$$\begin{aligned}L^i(q_0^{i'}, q_1^{i'}) =& \frac{(1 - P_1^i)^2(1 - q_0^i - q_1^i)^2}{\lambda_1^i \lambda_0^i}\\ =& L^i(q_0^i, q_1^i).\end{aligned}$$

Thus, $L^i(q_0^i, q_1^i)$ is symmetric about $(0.5, 0.5)$, which means for every point (q_0^i, q_1^i) on the left of $q_0^i + q_1^i = 1$, we can find a point on the right side of $q_0^i + q_1^i = 1$ that result in a same $L^i(q_0^i, q_1^i)$ value.

In terms of the constraints, first derive $F_1^i(q_0^i, q_1^i)$, $F_2^i(q_0^i, q_1^i)$, $F_3^i(q_0^i, q_1^i)$, $F_4^i(q_0^i, q_1^i)$ as:

$$\begin{aligned}F_1^i(q_0^i, q_1^i) =& \frac{Pr(X = 1)}{Pr(X = 1|Y = 0)} = \frac{\lambda_0^i}{q_1^i};\\ F_2^i(q_0^i, q_1^i) =& \frac{Pr(X = 1)}{Pr(X = 1|Y = 1)} = \frac{\lambda_1^i}{1 - q_1^i};\\ F_3^i(q_0^i, q_1^i) =& \frac{Pr(X = 0)}{Pr(X = 0|Y = 0)} = \frac{\lambda_0^i}{1 - q_0^i};\\ F_4^i(q_0^i, q_1^i) =& \frac{Pr(X = 0)}{Pr(X = 0|Y = 1)} = \frac{\lambda_1^i}{q_0^i}.\end{aligned} \tag{31}$$

if taking $q_0^{i'} = 1 - q_0^i$, $q_1^{i'} = 1 - q_1^i$ into $F_1^i(q_0^i, q_1^i)$, $F_2^i(q_0^i, q_1^i)$, $F_3^i(q_0^i, q_1^i)$ and $F_4^i(q_0^i, q_1^i)$:
$F_1^i(q_0^{i'}, q_1^{i'}) = \frac{\lambda_1^i}{1 - q_1^i} = F_2^i(q_0^i, q_1^i)$, $F_2^i(q_0^{i'}, q_1^{i'}) =$

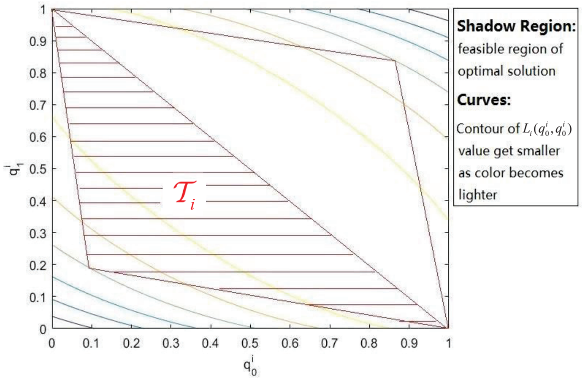


Fig. 9. Illustration of $L_i(q_0^i, q_1^i)$ and Region $\mathcal{T}_i$

$\frac{\lambda_0^i}{q_1^i} = F_1^i(q_0^i, q_1^i)$, $F_3^i(q_0^{i'}, q_1^{i'}) = \frac{\lambda_1^i}{q_0^i} = F_4^i(q_0^i, q_1^i)$, $F_4^i(q_0^{i'}, q_1^{i'}) = \frac{\lambda_0^i}{1-q_0^i} = F_1^i(q_0^i, q_1^i)$.
As a result, the four constraints functions are symmetric about point $(0.5, 0.5)$. Assume that $F_j^i(q_0^i, q_1^i)$, $j = 1, 2, 3, 4$ form a feasible region $\mathcal{T}_i$ for (q_0^i, q_1^i), thus for any point in this region, another point can be found on the other side of $q_0^i + q_1^i = 1$ also in $\mathcal{T}_i$. Now $(q_0^{i*}, q_1^{i*}) \in \mathcal{T}_i = \mathcal{T}_i \cap \{q_0^i + q_1^i \leq 1\}$.
Fig. 2 illustrates the function of $L_i(q_0^i, q_1^i)$ and feasible region $\mathcal{R}^i$ on the plane of q_0^i, q_1^i, curves in the figure is the contour line of $L_i(q_0^i, q_1^i)$, while the shadow region is the feasible region of $\mathcal{T}_i$ when ϵ is fixed.

3) Step 3
Since

$$\begin{aligned} F_1^i - F_2^i =& P_1^i + \frac{(1-P_1^i)(1-q_0^i)}{q_1^i} - (P_1^i + \frac{(1-P_1^i)q_0^i}{1-q_1^i}) \\ =& (1-P_1^i)\frac{1-q_0^i-q_1^i}{q_1^i(1-q_1^i)} \end{aligned}$$

and

$$\begin{aligned} F_4^i - F_3^i =& 1 - P_1^i + \frac{P_1^i(1-q_1^i)}{q_0^i} - (1 - P_1^i + \frac{P_1^i q_1^i}{1-q_0^i}) \\ =& (1-P_1^i)\frac{1-q_0^i-q_1^i}{q_1^i(1-q_1^i)}, \end{aligned}$$

for any point $(q_0^i, q_1^i) \in \mathcal{T}_i$, there is $F_1^i \geq F_2^i$ and $F_4^i \geq F_3^i$

$$\begin{aligned} \mathcal{T}_i = \{F_1^i \leq e^\epsilon\} \cap \{e^{-\epsilon} \leq F_2^i\} \cap \{e^{-\epsilon} \leq F_3^i\} \\ \cap \{F_4^i \leq e^\epsilon\} \cap \{q_0^i + q_1^i \leq 1\} \end{aligned} \tag{32}$$

From (32), assume that q_0^i is fixed to be q_0, then

$$q_1^i \geq \begin{cases} \frac{(1-P_1^i)(1-q_0)}{e^\epsilon - P_1^i} & if \quad q_0 \geq P_1^i/e^\epsilon \\ 1 - \frac{(e^\epsilon + P_1^i - 1)(q_0)}{P_1^i} & if \quad q_0 < P_1^i/e^\epsilon. \end{cases}$$

Since $L^i(q_0^i, q_1^i)$ is monotonically decreasing with q_0^i and q_1^i in $\mathcal{T}_i$, For any q_0^i in $\mathcal{T}_i$, its according optimal q_1^{i*} is the smallest value s.t. (q_0^i, q_1^{i*}) is in $\mathcal{T}_i$. It is the same with q_0^{i*}. On the other hand, F_1^i, F_2^i, F_3^i, F_4^i are all liner, their bounds value can be achieved when they are equal to their constraints values e^ϵ or $e^{-\epsilon}$. As a result,

$$(q_0^{i*}, q_1^{i*}) \in \begin{cases} \{F_1^i = e^\epsilon\} & if \quad q_0 \geq P_1^i/e^\epsilon \\ \{F_4^i = e^\epsilon\} & if \quad q_0 < P_1^i/e^\epsilon. \end{cases}$$

4) Step 4
To find the minimal value of $L^i(q_0^i, q_1^i)$, where $(q_0^i, q_1^i) \in \mathcal{T}_i$, the last thing we need to do is to test its monotocity of $L^i(q_0^i, q_1^i)$ given $F_1^i = e^\epsilon$ or $F_4^i = e^\epsilon$.

$$\begin{aligned} L^i_{F_1^i = e^\epsilon}(q_0^i, q_1^i) =& \frac{(1-P_1^i)^2(q_0^i + q_1^i - 1)^2}{\lambda_0^i \lambda_1^i} \\ =& \frac{[(1-P_1^i)(e^\epsilon - 1)q_1^i]^2}{\lambda_0^i \lambda_1^i}, \end{aligned}$$

To find the minimal value, taking derivative over $L^i_{F_1^i = e^\epsilon}$

$$\begin{aligned} \frac{\partial L^i_{F_1^i = e^\epsilon}(q_0^i, q_1^i)}{\partial q_1^i} =& (1-P_1^i)^2(e^\epsilon - 1)^2(\frac{{q_1^i}^2}{\lambda_0^i \lambda_1^i})' \\ =& (1-P_1^i)^2(e^\epsilon - 1)^2 q_1^i \frac{\lambda_1^i(\lambda_0^i - P_1^i q_1^i) + \lambda_0^i \lambda_1^i + P_1^i q_1^i \lambda_0^i}{(\lambda_0^i \lambda_1^i)^2} \end{aligned}$$

Which is obviously greater than 0, similarly, when $F_4^i = e^\epsilon$ is given:

$$\begin{aligned} \frac{\partial L^i_{F_4^i = e^\epsilon}(q_0^i, q_1^i)}{\partial q_0^i} =& {P_1^i}^2(e^\epsilon - 1)^2(\frac{{q_0^i}^2}{\lambda_0^i \lambda_1^i})' \\ =& {P_1^i}^2(e^\epsilon - 1)^2 q_0^i \frac{\lambda_0^i(\lambda_1^i - (1-P_1^i)q_0^i) + \lambda_1^i(1-P_1^i)q_0^i}{(\lambda_0^i \lambda_1^i)^2} \end{aligned}$$

Which is also greater then 0. Noticed that, when $q_0^i \leq \frac{P_1^i}{e^\epsilon}$, optimal point is $(q_0^{i*}, q_1^{i*}) = (\frac{P_1^i}{e^\epsilon}, \frac{1-P_1^i}{e^\epsilon})$, when $q_0^i > \frac{P_1^i}{e^\epsilon}$, the optimal point is also $(q_0^{i*}, q_1^{i*}) = (\frac{P_1^i}{e^\epsilon}, \frac{1-P_1^i}{e^\epsilon})$. Thus the global optimal solution is $(\frac{P_1^i}{e^\epsilon}, \frac{1-P_1^i}{e^\epsilon})$.

□

Appendix B
Proof of Theorem 5

Proof. We first drive the formulation of $Var(\hat{S}')$.

$$\begin{aligned} Var(\hat{S}_i') =& (\frac{(1-q^i) \cdot P_1^i}{q^i \circledast P_1^i} - \frac{q^i \cdot P_1^i}{1 - q^i \circledast P_1^i})^2 Var(Y_i) \\ =& {P_1^i}^2 \frac{[(2q^i - 1)(P_1^i - 1)]^2}{(q^i \circledast P_1^i)(1 - q^i \circledast P_1^i)}. \end{aligned} \tag{33}$$

Combine (33) and (8)

$$\begin{aligned} &\mathcal{E}_{Sym}(q^i) = E[(S - \hat{S}_2)^2] = \\ &\sum_{i=1}^{N} [P_1^i(1-P_1^i) - {P_1^i}^2 \frac{[(2q^i - 1)(P_1^i - 1)]^2}{(q^i \circledast P_1^i)(1 - q^i \circledast P_1^i)}]. \end{aligned} \tag{34}$$

Now, consider privacy constraints. When $P_1^i \leq 0.5$: $F_1^i(0.5) = F_4^i(0.5)$, and $F_1^i(q^i) > F_4^i(q^i)$ when $q^i < 0.5$; $F_1^i(q^i) < F_4^i(q^i)$ when $q^i > 0.5$. Define the lower bound of F_1^i and F_4^i as $P^i_{L(1,4)}$; upper bound of F_1^i and F_4^i as $P^i_{U(1,4)}$. Obviously, $P^i_{L(1,4)} = \max\{{F_1^i}^{-1}(e^\epsilon), {F_4^i}^{-1}(e^\epsilon)\} = {F_1^i}^{-1}(e^\epsilon)$.

On the other hand, the Upper bound of F_1^i and F_4^i depends on the value of ϵ and P_1^i, that is:

$$\begin{cases} P^i_{U(1,4)} = 1 & e^{-\epsilon} \le P^i_1 \\ P^i_{U(1,4)} = F_1^{i^{-1}}(e^{-\epsilon}) & e^{-\epsilon} > P^i_1. \end{cases}$$

On the other hand, $F_2^i(0.5) = F_3^i(0.5)$, and $F_2^i(q^i) < F_3^i(q^i)$ when $q^i < 0.5$; $F_2^i(q^i) < F_3^i(q^i)$ when $q^i > 0.5$. Define the lower bound of F_2^i and F_3^i as $P^i_{L(2,3)}$; upper bound of F_2^i and F_3^i as $P^i_{U(2,3)}$.

Obviously, $P^i_{U(2,3)} = \max\{F_2^{i^{-1}}(e^{\epsilon}), F_3^{i^{-1}}(e^{\epsilon})\} = F_2^{i^{-1}}(e^{\epsilon})$. In terms of the Lower bound, it is similar to the case of F_1^i and F_4^i:

$$\begin{cases} P^i_{L(2,3)} = 0 & e^{-\epsilon} \le P^i_1 \\ P^i_{L(2,3)} = F_2^{i^{-1}}(e^{-\epsilon}) & e^{-\epsilon} > P^i_1. \end{cases}$$

Next, we can compare the two set of lower bounds and upper bounds to obtain the global solution.

Compare $P^i_{L(1,4)}$ with $P^i_{L(2,3)}$:

When $e^{-\epsilon} > P^i_1$:

$$\begin{aligned} P^i_{L(1,4)} - P^i_{L(2,3)} &= F_1^{i^{-1}}(e^{\epsilon}) - F_2^{i^{-1}}(e^{-\epsilon}) \\ &= \frac{1 - P^i_1}{e^{\epsilon} - 2P^i_1 + 1} - \frac{e^{-\epsilon} - P^i_1}{e^{-\epsilon} - 2P^i_1 + 1} \\ &= \frac{P^i_1(e^{\epsilon} + e^{-\epsilon} - 2)}{(e^{\epsilon} - 2P^i_1 + 1)(e^{-\epsilon} - 2P^i_1 + 1)} \ge 0; \end{aligned}$$

When $e^{-\epsilon} \le P^i_1$: $P^i_{L(1,4)} > P^i_{L(2,3)} = 0$. So, $P^i_{L(1,2,3,4)} = P^i_{L(1,4)} = F_1^{i^{-1}}(e^{\epsilon})$, for $P^i_1 \le 0.5$. We can use the same way to check that: $P^i_{U(1,2,3,4)} = P^i_{U(2,3)} = F_2^{i^{-1}}(e^{\epsilon})$, for $P^i_1 \le 0.5$. It is the same to find the results when $P^i_1 > 0.5$: $P^i_{L(1,2,3,4)} = P^i_{L(1,4)} = F_4^{i^{-1}}(e^{\epsilon})$, for $P^i_1 > 0.5$; $P^i_{U(1,2,3,4)} = P^i_{U(2,3)} = F_3^{i^{-1}}(e^{\epsilon})$, for $P^i_1 > 0.5$.

On the other hand, we also want to know which bound value may result in the minimal value of MSE, it is easy to do so by comparing the two resulting MSE values:

$$F_1^{i^{-1}}(e^{\epsilon}) + F_2^{i^{-1}}(e^{\epsilon}) = \frac{1 - P^i_1}{e^{\epsilon} - 2P^i_1 + 1} + \frac{e^{\epsilon} - P^i_1}{e^{\epsilon} - 2P^i_1 + 1} = 1.$$

Thus, $F_1^{i^{-1}}(e^{\epsilon})$ and $F_2^{i^{-1}}(e^{\epsilon})$ are axisymmetric about 0.5. We can also check that $F_3^{i^{-1}}(e^{\epsilon}) + F_4^{i^{-1}}(e^{\epsilon}) = 1$ and $F_3^{i^{-1}}(e^{\epsilon})$ and $F_4^{i^{-1}}(e^{\epsilon})$ are also axisymmetric about 0.5.

In conclusion: $\mathcal{E}_{Sym}(F_1^{i^{-1}}(e^{\epsilon})) = \mathcal{E}_{Sym}(F_2^{i^{-1}}(e^{\epsilon}))$; $\mathcal{E}_{Sym}(F_3^{i^{-1}}(e^{\epsilon})) = \mathcal{E}_{Sym}(F_4^{i^{-1}}(e^{\epsilon}))$. □

References

[1] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," tech. rep., 1998.

[2] C. Dwork, "Differential privacy: A survey of results," in *Theory and Applications of Models of Computation: 5th International Conference, TAMC* (M. Agrawal, D. Du, and Z. Duan, eds.), pp. 1–19, 2008.

[3] C. Dwork, "Differential privacy," in *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Part II* (M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds.), pp. 1–12, 2006.

[4] C. Dwork, F. McSherry, and K. Nissim, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography: Third Theory of Cryptography Conference*, pp. 265–284, 2006.

[5] A. FITZPATRICK, "What to do after the massive yahoo hack," 2016.

[6] R. Chen, H. Li, A. K. Qin, S. P. Kasiviswanathan, and H. Jin, "Private spatial data aggregation in the local setting," in *2016 IEEE 32nd ICDE*, pp. 289–300, May 2016.

[7] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[8] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," in *Advances in Neural Information Processing Systems 27*, pp. 2879–2887, Curran Associates, Inc., 2014.

[9] A. D. Sarwate and L. Sankar, "A rate-disortion perspective on local differential privacy," in *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 903–908, Sept 2014.

[10] S. Xiong, A. D. Sarwate, and N. B. Mandayam, "Randomized requantization with local differential privacy," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2189–2193, March 2016.

[11] lfar Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 21st ACM CCCS*, 2014.

[12] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in apple's implementation of differential privacy on MacOS 10.12," *CoRR*, vol. abs/1709.02753, 2017.

[13] T. Wang, J. Blocki, N. Li, and S. Jha, "Locally differentially private protocols for frequency estimation," in *26th USENIX Security 17*, pp. 729–745, USENIX Association, 2017.

[14] T.-H. H. Chan, E. Shi, and D. Song, "Optimal lower bound for differentially private multi-party aggregation," in *Proceedings of the 20th Annual ECA*, ESA'12, pp. 277–288, 2012.

[15] R. Bassily, K. Nissim, U. Stemmer, and A. Thakurta, "Practical locally private heavy hitters," *CoRR*, vol. abs/1707.04982, 2017.

[16] C. Huang, P. Kairouz, X. Chen, L. Sankar, and R. Rajagopal, "Context-aware generative adversarial privacy," *CoRR*, vol. abs/1710.09549, 2017.

[17] W. Wang, L. Ying, and J. Zhang, "On the tradeoff between privacy and distortion in differential privacy," *CoRR*, vol. abs/1402.3757, 2014.

[18] S. Asoodeh, F. Alajaji, and T. Linder, "Notes on information-theoretic privacy," in *2014 52nd Allerton*, pp. 1272–1278, Sept 2014.

[19] G. Wu and X. Xia, "Extending differential privacy for treating dependent records via information theory," *CoRR*, vol. abs/1703.07474, 2017.

[20] W. Wang, L. Ying, and J. Zhang, "On the relation between identifiability, differential privacy, and mutual-information privacy," *IEEE Transactions on Information Theory*, vol. 62, pp. 5018–5029, Sept 2016.

[21] S. Asoodeh, F. Alajaji, and T. Linder, "On maximal correlation, mutual information and data privacy," in *IEEE CWIT*, pp. 27–31, July 2015.

[22] M. E. Andrés, N. E. Bordenabe, and K. Chatzikokolakis, "Geo-indistinguishability: Differential privacy for location-based systems," *CoRR*, vol. abs/1212.1984, 2012.

[23] Y. Cao, M. Yoshikawa, Y. Xiao, and L. Xiong, "Quantifying differential privacy under temporal correlations," in *2017 IEEE 33rd ICDE*, pp. 821–832, April 2017.

[24] F. Tramèr and Z. Huang, "Differential privacy with bounded priors: Reconciling utility and privacy in genome-wide association studies," in *Proceedings of the 22Nd ACM SIGSAC*, pp. 1286–1297, 2015.

[25] F. Calmon and N. Fawaz, "Privacy against statistical inference," 10 2012.

[26] F. A. S. Asoodeh and T. Linder, "Privacy-aware mmse estimation," in *2016 IEEE ISIT*, pp. 1989–1993, July 2016.

[27] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*. McGraw-Hill, 2002.

[28] Z. Qin, Y. Yang, and T. Yu, "Heavy hitter estimation over set-valued data with local differential privacy," in *Proceedings of the 2016 ACM SIGSAC*, CCS '16, pp. 192–203, 2016.

[29] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Commun. ACM*, vol. 13, pp. 422–426, July 1970.