Local Information Privacy with Bounded Prior

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*—A localized privacy protection notion: local information privacy (LIP) is studied in this paper. As a context-aware notion that considers prior knowledge, the LIP notion is shown to provide increased utility than local differential privacy (LDP). Within the scope of LIP, we further consider scenarios with uncertainty on the prior knowledge, i.e., the prior is bounded within a certain range or the prior is arbitrary. The former case is defined as bounded-prior LIP (BP-LIP), and the latter as worst-case LIP (WC-LIP).**

The contributions of this paper are three-fold: We first provide theoretical results which show the connections of these new definitions with LDP; Secondly, we present an optimization framework for privacy-preserving data collection, with the goal of minimizing the expected squared error while satisfying BP-LIP and WC-LIP privacy constraints. Utility-privacy tradeoffs are obtained in closed-form. At last, we validate our conclusions by numerical analysis and real-world data simulation. Our results show that the notion of bounded-prior LIP can achieve better utility-privacy tradeoff compared to context free notion of LDP.

I. Introduction

Privacy-preserving data collection methods have various real-world applications, such as data analytics [1], frequency estimation [2], and itemset mining [3]. In the data privacy research community, differential privacy (DP) [4] has been accepted as the *de facto* standard as it provides a rigours privacy guarantee which is measured explicitly by the privacy budget ϵ. There are two widely studied notions of DP: centralized DP, which relies on a trusted server that manipulates the perturbation and gives noisy output for certain queries; and local DP (LDP) [5], in which users are allowed to directly publish perturbed data to untrusted curator. "Localized" means the data perturbation is run at the user rather than at the server. Nowadays, The increasing popularity of LDP can be attributed to two factors: a) first, a trusted server may not always exist in certain application scenarios; and even if a curator is supposedly trusted, various privacy breaches in practice could occur due to internal compromises or other reasons; and b) aggregated data from multiple users may itself limit local utility that an individual user can provide. In fact, most individuals' data are collected directly, and hence, LDP notion has led to various privacy preserving mechanisms, such as [6]–[8], and Google's RAPPOR (randomized aggregatable privacy-preserving ordinal response) [9]. Even though LDP based mechanisms have already been widely applied, either they provide limited utility or the privacy is not well protected. For instance, Tang *et al.* show that, in Apple's LDP mechanism [10], the privacy budget is too large to provide any useful privacy protection [11]. In 2017, Wang *et al.* provide a variety of LDP protocols for frequency estimation [2] and compare their performance with Google's RAPPOR. However, including RAPPOR, none of these protocols provides considerable estimation accuracy under a reasonable small ϵ.

The main reason why it is hard for LDP to achieve an ideal utility-privacy tradeoff is because the definition provides a rigours privacy guarantee against worst-case (context-free) adversary that can possess arbitrary background knowledge. A promising way to increase utility beyond the definition of LDP is to introduce prior information into the picture, which allows an explicit measurement over the adversary's ability. The prior knowledge can be obtained through many ways in reality, such as the theoretical research results can be used as the prior knowledge for clinical treatment; past survey results can be used as prior for future data gathering; the classification from neural network trained by previous data can be used as prior for prediction. The privacy notions considering prior knowledge are usually referred to as "context-aware" notions. Information theoretic privacy notions [1], [12], [13] fall into this category, such as mutual information privacy (MIP) [14], which limits the average information leakage between the raw data and the perturbed data. However, MIP is a relatively weak notion compared to LDP, as mutual-information is the expected KL (Kullback Leibler) divergence between prior and posterior (where expectation is over all outcomes of data). In [15], the notion of local information privacy (LIP) is presented, which bounds on the ratio between the prior and posterior of the data after taking observations. Since the privacy notion is tailored to the specific prior knowledge adversary has, the privacy guarantee can be viewed as relaxed, and data perturbation can be done in a prior related manner to increase the data utility.

We next use a toy example of concrete mechanisms to recap and reinforce the intuition behind introducing prior.

Example 1. *Consider a one to one (one curator and one user) privacy preserving binary data collection model, where user's data is denoted as X, which takes 0 or 1 with certain probability $P_1 = Pr(X = 1)$. The raw data X is randomly perturbed to an output Y, with certain probabilities, the set of which is called perturbation parameter of the mechanism and is denoted as q (context-free) and q_0, q_1 (context-aware) in Fig. 1. This model allows the curator to make an estimation*

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

978-1-5386-8088-9/19/$31.00 ©2019 IEEE

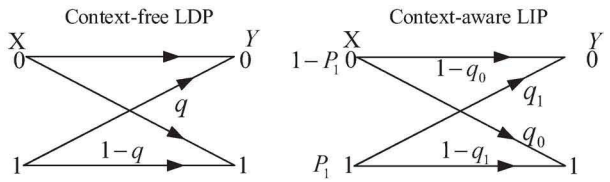

Fig. 1. An illustration of binary randomize response perturbation mechanism.

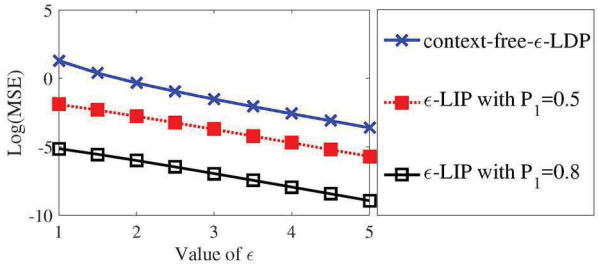

Fig. 2. The comparison of MSEs under different privacy notions and priors.

on X by taking observations on Y.

The model described above is usually referred to as a randomized response perturbation mechanism. To achieve an ideal accuracy, no matter what estimator the curator uses, the desired probabilities to perturb: q, q_1 and q_0 should be designed as small as possible, however, they can not be too small in order to satisfy the privacy constraints. Based on different privacy definitions, the minimal requirements of the perturbation probabilities are different. For the context-free LDP model, likelihood ratio is bounded by e^ϵ, and we have $\min q = \frac{1}{1+e^\epsilon}$. The context-aware LIP model, on the other hand, requires that the ratio between prior and posterior is bounded by e^ϵ, thus q_0 and q_1 are designed according to the priors: $\min q_0 = Pr(Y=1|X=0) = \frac{P_1}{e^\epsilon}$, $\min q_1 = Pr(Y=0|X=1) = \frac{1-P_1}{e^\epsilon}$ [15]. After observing Y, the curator is able to make a further estimation on X. The estimator of LDP: $\hat{X}_{LDP}$ is a function of q: $\hat{X}_{LDP} = (1-q)Y + \frac{q}{2}$ [2]; while in the LIP model, $\hat{X}_{LIP}$ is designed with q_0, q_1 as a function of P_1: $\hat{X}_{LIP} = (1 - \frac{1-P_1}{e^\epsilon})Y + \frac{P_1}{e^\epsilon}(1-Y)$ [15].

Fig. 2 shows the mean square errors of the estimation under three cases: "Context-free", "$P_1 = 0.5$" and "$P_1 = 0.8$". We can observe an obvious increment in the accuracy of the case when $P_1 = 0.8$ comparing with other two cases, which shows the usefulness of prior knowledge. Notice that when $P_1 = 0.5$, the prior knowledge does not help for data collecting, however, there's still an increasing gap between LIP when $P_1 = 0.5$ and the LDP. This is because the privacy guarantee is relatively relaxed than the LDP (for more details, see Sec. II).

While the context-aware LIP provides increased utility than LDP, it can be difficult to implement in real-world applications, as it assumes that the exact prior knowledge is available for both the users and curator. In the above example, $P_1 = 0.5$ or $P_1 = 0.8$ is a common knowledge so that user and curator can design prior related mechanism. However, more often than not, the prior knowledge exists but with uncertainty. We study this uncertain prior LIP notion, and present two other LIP notions with bounded prior knowledge and arbitrary prior knowledge. Then we develop explicit connections to LDP.

The main contributions of this paper are three-fold:

(1) Motivated by the above discussion, we propose and study two variations of LIP: worst-case LIP (WC-LIP) and bounded-prior LIP (BP-LIP). We formally show the relationships between them and LDP: ϵ-WC-LIP is equivalent to ϵ-LDP; ϵ-BP-LIP is sandwiched between 2ϵ-LDP and ϵ-LDP as a function of the uncertainty of the prior knowledge.

(2) We study a binary data perturbation and estimation model which achieves optimality for single user collection and multiple users data aggregation simultaneously. This binary model has applications in privacy-preserving survey and frequency estimation problems. We formulate optimization problems to minimize the mean square of estimation error subject to proposed privacy constraints. The perturbation parameters are derived in closed-form.

(3) We use both the theoretical results and simulations to compare with the utility privacy tradeoffs between BP-LIP, WC-LIP notions and LDP notions, and show the advantages of BP-LIP and WC-LIP notions.

II. Privacy Definitions

Next, we recap the related localized privacy definitions, including LDP and LIP, and briefly discuss the intuitions behind each notion. Then we introduce definitions of the proposed variants of LIP, including WC-LIP and BP-LIP, and investigate the relationships between them.

Local Differential Privacy: As a localized version of differential privacy, LDP provides context-free privacy guarantee against the worst-case adversary. It requires any two different inputs within the data range have similar probabilities (measured by ϵ) to result in a same output:

Definition 1. *(ϵ-Local Differential Privacy (LDP)) [5] 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 $\forall y \in Range(\mathcal{M})$:*

$$\frac{Pr(Y=y|X=x)}{Pr(Y=y|X=x')} \leq e^\epsilon. \tag{1}$$

The definition of LDP does not allow the mechanism to depend on knowledge of prior. To explicitly introduce prior knowledge, Local information privacy is proposed in [15].

Local Information Privacy: We first define the distribution over a dataset: Let $\mathbb{D} = \{1, 2, ..., |\mathbb{D}|\}$, and define $\mathbf{P} = \{P_1, P_2, ..., P_{|\mathbb{D}|}\}$ as a prior distribution over the dataset $\mathbb{D}$, i.e., $P_x = Pr(X = x)$, where $x \in \mathbb{D}$. For a given distribution $\mathbf{P}$, the notion of LIP is defined by bounding the ratio between the prior and the posterior, which guarantees that after observing the output, an adversary cannot infer too much additional information other than its prior.

Definition 2. *(ϵ-Local Information Privacy (LIP)) [15] A mechanism $\mathcal{M}$ which takes input X with prior distribution $\mathbf{P}$ and outputs Y satisfies ϵ-LIP for some $\epsilon \in \mathbb{R}^+$, if $\forall x, y \in \mathbb{D}$:*

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

In [15], it is shown that LIP is a weaker notion than LDP. In particular, if a mechanism $\mathcal{M}$ satisfies ϵ-LDP, it also satisfies ϵ-LIP; If $\mathcal{M}$ satisfies ϵ-LIP, it satisfies 2ϵ-LDP.

The notion of LIP is context-aware and allows for designing perturbation mechanisms that provide higher utility than LDP by leveraging the prior. However, the assumption that the curator and the users possess the exact prior knowledge prohibits this notion from real world applications, since, the available knowledge about the prior may not be accurate. Depending on different amount of knowledge about the prior, we propose two new definitions.

Worst-Case Local Information Privacy: Worst-case LIP means that the privacy definition (2) must be satisfied for all possible priors, which is defined as $R_{wc} = \{\mathbf{P} : P_x \in [0,1], \sum_{x\in\mathbb{D}} P_x = 1\}$, meaning that in the $R^{|\mathbb{D}|}$ space, the distribution $\mathbf{P}$ can take any point on the plane of $\sum_{x\in\mathbb{D}} P_x = 1$.

Definition 3. *(ϵ-Worst-Case Local Information Privacy (WC-LIP)) A mechanism $\mathcal{M}$ which takes input X and outputs Y satisfies ϵ-WC-LIP for some $\epsilon \in \mathbb{R}^+$, if $\forall x, y \in \mathbb{D}$, $\forall \mathbf{P} \in R_{wc}$, (2) is satisfied.*

Next, we show WC-LIP is equivalent to LDP.

Theorem 1. *A mechanism satisfies ϵ-WC-LIP if and only if it satisfies ϵ-LDP.*

Proof. To prove the "if" part, consider a $\mathcal{M}$ that takes any two inputs $X = x$, $X = x'$ and outputs the same $Y = y$.

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:

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

By switching inputs, we can also get:

$$Pr(Y = y) \geq e^{-\epsilon} Pr(Y = y|X = x), \forall P_x \in [0,1].$$

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

For the "only if" part, if $\mathcal{M}$ satisfies ϵ-WC-LIP, $\forall \mathbf{P} \in R_{wc}$, by Bayes rule and looking at one side:

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

As $P_x \in \mathbf{P}$ can take any value from 0 to 1 in the worst-case. Thus when $P'_x = 1$, $P_x = 0, \forall x \neq x'$, there is:

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

which means the definition of LDP is also satisfied. □

The equivalence of these two definitions means that in a worst-case setting, the LIP privacy notion (if satisfied among all possible priors), is equivalent to LDP.

Bounded-Prior Local Information Privacy: On the other hand, in reality, the common prior knowledge available for user and the adversary usually lies within a range. For example, if it is a common knowledge that something is more likely to happen than not. We know, $0.5 \leq P \leq 1$, where P is the prior to happen. Similarly, we define the feasible region of the bounded prior as R_{bp}, which is a subset of R_{wc}: $R_{bp} = \{\mathbf{P} : P_x \in [a_x, b_x], \sum_{x\in\mathbb{D}} P_x = 1\}$. R_{bp} defines a convex region on the plane of $\sum_{x\in\mathbb{D}} P_x = 1$ in $R^{|\mathbb{D}|}$.

Definition 4. *(ϵ-Bounded Prior Local Information Privacy (BP-LIP)) A mechanism $\mathcal{M}$ which takes input X and outputs Y satisfies ϵ-BP-LIP for some $\epsilon \in \mathbb{R}^+$, if $\forall x, y \in \mathbb{D}$, $\forall \mathbf{P} \in R_{bp}$, (2) is satisfied.*

The next theorem states the relationship between BP-LIP and LDP:

Theorem 2. *If a mechanism $\mathcal{M}$ satisfies ϵ-BP-LIP, then the privacy strength of $\mathcal{M}$ is sandwiched by ϵ-LDP and 2ϵ-LDP.*

Proof. The first part means ϵ-LDP implies ϵ-BP-LIP: By Theorem 1, we know ϵ-LDP is equivalent to ϵ-WC-LIP, and ϵ-WC-LIP implies ϵ-BP-LIP, as R_{bp} is a subset of R_{wc}.

The second part means ϵ-BP-LIP implies 2ϵ-LDP: from [15], we know ϵ-LIP implies 2ϵ-LDP, where ϵ-LIP can be viewed as a special case of ϵ-BP-LIP (any possible $\mathbf{P}$ belongs to R_{bp}), Thus, ϵ-BP-LIP implies ϵ-LIP which further implies ϵ-LDP. □

We next investigate how R_{bp} influences the relationship between ϵ-BP-LIP and ϵ-LDP under a binary model: $\mathbb{D} = \{0, 1\}$, $P_1 = Pr(X = 1)$. Define the feasible region R_{bp} as $P_1 \in [a, b]$, where $0 \leq a \leq b \leq 1$. When the prior of P_1 is fixed, $Pr(X = 0)$ is also fixed. Thus $P_1 \in [a, b]$ is sufficient to represent R_{bp}. By Definition 4, if $\mathcal{M}$ satisfies ϵ-BP-LIP, then $\forall P_1 \in [a, b]$, (2) is satisfied. Suppose $\mathcal{M}$ also satisfies LDP for some $\epsilon' \geq 0$, for a fixed $[a, b]$, we have:

Theorem 3. *In the binary model, if $\mathcal{M}$ satisfies ϵ-BP-LIP with $P_1 \in [a, b]$, it also satisfies ϵ'-LDP, where ϵ' corresponds to the following values under each case:*

When $a + b \leq 1$:
if $\epsilon \leq \ln \frac{1-b}{a}$: $\epsilon' = \ln \frac{1-a}{e^{-\epsilon}-a}$; otherwise, $\epsilon' = \ln \frac{e^{\epsilon}+b-1}{b}$;
When $a + b > 1$:
if $\epsilon \leq \ln \frac{a}{1-b}$: $\epsilon' = \ln \frac{b}{e^{-\epsilon}-1+b}$; otherwise, $\epsilon' = \ln \frac{e^{\epsilon}-a}{1-a}$.

Proof. Detailed proof is shown in Appendix A. □

It can be readily shown for all the cases, ϵ' is sandwiched between ϵ and 2ϵ. Also notice that as a function of a and b, ϵ' is monotonically increasing with a, and monotonically decreasing with b. Which means that the maximum ϵ' is achieved when a goes to b, where BP-LIP is equivalent to original LIP; The minimum value of ϵ' is achieved when a goes to 0 and b goes to 1, where BP-LIP is equivalent to WC-LIP. Generally speaking, BP-LIP can be viewed with adaptive privacy guarantee: when the distance between a and b is large, which means there's much uncertainty on the prior, ϵ' decreases, hence, the BP-LIP notion is strong to protect against any possible privacy inference. When the distance between a and b is small, which means the prior knowledge of the adversary is quite clear, ϵ' increases, hence, the BP-LIP notion is designed to defend against privacy attack with known priors.

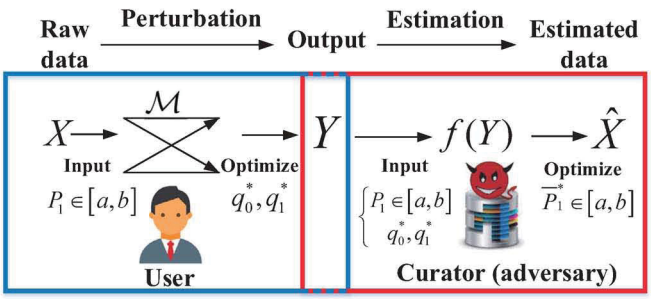

Fig. 3. Single user privacy-preserving data collection configuration.

III. PRIVACY-PRESERVING DATA COLLECTION MODEL

From [15], The local multiple users' data aggregation problem is optimized when each user's local data collecting problem is optimized if each user perturbs and publishes data independently. In this paper, we specifically investigate on single user data collection problem.

A. System and Threat Model

Consider an individual possesses a private answer to some query from a curator, which is denoted as X. It is assumed that X takes value from a binary domain $\mathbb{D} = \{0, 1\}$ with certain prior probabilities $P_1 = Pr(X = 1)$, $1 - P_1 = Pr(X = 0)$. It is assumed that the exact value of P_1 is not available for the individual and curator, however, both of them know that P_1 is bounded within $[a, b]$, where $0 \leq a \leq b \leq 1$. In order to use P_1 as a parameter to estimate, the curator picks $\bar{P}_1$ from $[a, b]$ as the prior. The user, on the other hand, designs a privacy-preserving mechanism $\mathcal{M}$ to defend against the adversary with arbitrary prior within $[a, b]$. $\mathcal{M}$ takes input X and perturbs to Y with certain probabilities, the set of which is denoted as perturbation parameters: $q_0 = Pr(Y = 1|X = 0)$; $q_1 = Pr(Y = 0|X = 1)$. After receiving Y, the curator wants to estimate X, denote the estimation as $\hat{X} = f(Y)$, where $f()$ is an estimator, in this paper, we deploy the minimized mean square error (MMSE) estimator [16]: $f(Y) = E[X|Y]$, as the exact P_1 is unavailable, the estimator becomes $f(Y) = E_{\bar{P}_1}[\bar{X}|Y]$, where $\bar{X}$ is the random variable which takes the distribution of $\bar{P}_1$. The configuration is shown in Fig. 3.

It is assumed that the curator possesses the knowledge of the perturbation mechanism as well as the range of the prior $[a, b]$ (same with the user). The curator (adversary) is assumed to be untrusted for multiple reasons, such as the private data is profitable, or the curator is hacked. As a result, user wants to cooperate with the curator for an estimation while remain some uncertainty on the data. The accurate estimation is denoted as "Utility", and the uncertainty is denoted as "Privacy".

B. Problem Formulation

Utility: In this paper, utility is measured by the inverse of mean square error (MSE) between the raw data and the estimation, where the MSE can be expressed as:

$$\mathcal{E} = E_{P_1}[(X - E_{\bar{P}_1}[\bar{X}|Y])^2]. \tag{5}$$

Next we derive the privacy constraints for them.

Privacy: The privacy is measured by the BP-LIP, for the binary model, define $F_{xy} = \frac{Pr(X=x)}{Pr(X=x|Y=y)}$, then, $\forall P_1 \in [a, b]$:

$$e^{-\epsilon} \leq \{F_{10}, F_{01}, F_{11}, F_{00}\} \leq e^{\epsilon}. \tag{6}$$

Eq. (6) forms a feasible region for q_0 and q_1. User wants to help the curator to estimate while defending against attack with any possible P_1. So, the optimal perturbation parameters q_0^* and q_1^* are found at the optimal solutions of problem to minimize the MSE subject to the BP-LIP constraints:

$$\begin{aligned} &\min \mathcal{E}(q_0, q_1), \\ &s.t.\ (6),\ \forall \bar{P}_1 \in [a, b] \text{ and } \forall P_1 \in [a, b]. \end{aligned} \tag{7}$$

As the curator knows $[a, b]$ and user's mechanism, q_0^* and q_1^* are also available for him. To build the MMSE estimator, which is prior-related, he picks $\bar{P}_1$ from $[a, b]$. Notice that for any fixed $\bar{P}_1$, we can find a $P_1 \in [a, b]$, such that P_1 maximize $\mathcal{E}$, denote this MSE as the $\max_{P_1} \mathcal{E}$ under $\bar{P}_1$, then we want to find an optimal $\bar{P}_1^*$ under which the mechanism results in the minimum $\max_{P_1} \mathcal{E}$. Briefly speaking, the optimal $\bar{P}_1^*$ should guarantee that the MSE in the worst-case ($\forall P_1 \in [a, b]$) is not too large. As a result, $\bar{P}_1^*$ is at the solution of the optimization problem:

$$\min_{\bar{P}_1 \in [a,b]} \max_{P_1 \in [a,b]} \mathcal{E}(P_1, \bar{P}_1). \tag{8}$$

We next derive the optimal solutions of q_0^*, q_1^* and $\bar{P}_1^*$.

C. Main Results

The next theorem derives the optimal q_0 and q_1 for any $P_1, \bar{P}_1 \in [a, b]$.

Theorem 4. *For the optimization problem defined in* (7)*, the optimal* q_0 *and* q_1 *are* $q_0^* = \frac{b}{b-a+e^{\epsilon}}$ *and* $q_1^* = \frac{1-a}{b-a+e^{\epsilon}}$.

Proof. For an arbitrary but fixed $\bar{P}_1$, the estimator is:

$$E[\bar{X}|Y] = Pr(\bar{X} = 1|Y). \tag{9}$$

Thus receiving $Y = 1$, the curator estimates $\hat{X}_1 = Pr(\bar{X} = 1|Y = 1) = \frac{\bar{P}_1(1-q_1)}{Pr(Y=1)}$; When $Y = 0$, the curator estimates $\hat{X}_0 = Pr(\bar{X} = 1|Y = 0) = \frac{\bar{P}_1 q_1}{Pr(Y=0)}$.

The MSE function can be rewritten as:

$$\begin{aligned} &\sum_{x,y \in \{0,1\}} (x - E[\bar{X}|Y])^2 Pr(X = x) Pr(Y = y|X = x) \\ &= (1 - \hat{X}_1)^2 P_1 (1 - q_1) + (1 - \hat{X}_0)^2 P_1 q_1 \\ &+ (\hat{X}_1)^2 (1 - P_1) q_0 + (\hat{X}_0)^2 (1 - P_1)(1 - q_0). \end{aligned} \tag{10}$$

It can be readily checked by taking $q_1' = 1 - q_1$ and $q_0' = 1 - q_0$ that the function of $\mathcal{E}$ is symmetric about $(0.5, 0.5)$. Thus we can narrow the whole feasible region by adding a constraint: $q_1 + q_0 \leq 1$. Then comparing with $\hat{X}_1$ and $\hat{X}_0$, we have:

$$\hat{X}_1 - \hat{X}_0 = \frac{(1 - \bar{P}_1)(1 - q_0 - q_1)}{Pr(Y = 1)[1 - Pr(Y = 1)]} \geq 0. \tag{11}$$

Which means in the simplified region, we always have $\hat{X}_1 \geq \hat{X}_0$. Now consider the four terms in (10), we want to control q_1 and q_0 to minimize $\mathcal{E}$: the first term wants q_1 large while the second term wants it small; the fourth term want q_0 large while the third term wants it small. However, as $(1 - \hat{X}_1)^2 \leq$

$(1-\hat{X}_0)^2$, also $(\hat{X}_1)^2 \geq (\hat{X}_0)^2$. Which means in order to minimize $\mathcal{E}$, we minimize q_1 and q_0. The minimized q_1 and q_0 are found at the boundary of the privacy constraints, which is achieved when $\max \frac{Pr(Y=1)}{q_0} = e^\epsilon$ and $\max \frac{Pr(Y=0)}{q_1} = e^\epsilon$.

$$\max \frac{Pr(Y=1)}{q_0} = \max \frac{P_1(1-q_1)+(1-P_1)q_0}{q_0}. \tag{12}$$

As $1-q_1 \geq q_0$, $\max_{P_1 \in [a,b]} \frac{Pr(Y=1)}{q_0} = \frac{b(1-q_1)+(1-b)q_0}{q_0} = e^\epsilon$. Similarly, $\max_{P_1 \in [a,b]} \frac{Pr(Y=0)}{q_1} = \frac{aq_1+(1-a)(1-q_0)}{q_1} = e^\epsilon$. The optimal solutions are calculated accordingly. □

Intuitively, to increase utility, we need the probability of perturbation as small as possible (when $q_0 + q_1 \leq 1$), 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.

Observing the expression of q_0^* and q_1^*, when $a = b = P_1$, which means the prior knowledge is certain and fixed, in this case $q_0^* = \frac{P_1}{e^\epsilon}$ and $q_1^* = \frac{1-P_1}{e^\epsilon}$ which are same with the optimal solutions of LIP [15]; When $a = 0$, $b = 1$, we have the optimal solutions for the WC-LIP: $q_0^* = q_1^* = \frac{1}{1+e^\epsilon}$, which is independent of prior and is identical to the optimal solutions of the LDP [2]. This result show that the BP-LIP connects the notions of LIP, WC-LIP and LDP together by adjusting prior uncertainty. Also, interestingly, the result implies that the WC-LIP and LDP are equivalent even in optimal mechanism.

The next step is finding the optimal $\bar{P}_1$, which can be derived by the following lemma.

Lemma 1. *If $\bar{P}_1 \neq P_1$, $\mathcal{E}(\bar{P}_1) > \mathcal{E}(P_1)$ and:*

$$\mathcal{E}(P_1) - \mathcal{E}(\bar{P}_1) = -\frac{(e^\epsilon - 1 + b)b(e^\epsilon - 1)^2}{(e^\epsilon - a)(1-a)e^{3\epsilon}}(P_1 - \bar{P}_1)^2. \tag{13}$$

Proof. As the optimal perturbation parameters are $q_0^* = \frac{b}{b-a+e^\epsilon}$ and $q_1^* = \frac{1-a}{b-a+e^\epsilon}$, the estimator becomes:

$$\begin{aligned} \hat{X}_1(\bar{P}_1) &= E[\bar{X}|Y=1] = \frac{\bar{P}_1(b+e^\epsilon-1)}{b-\bar{P}_1+\bar{P}_1 e^\epsilon}; \\ \hat{X}_0(\bar{P}_1) &= E[\bar{X}|Y=0] = \frac{\bar{P}_1(1-a)}{\bar{P}_1-a-e^\epsilon-\bar{P}_1 e^\epsilon}. \end{aligned} \tag{14}$$

By (10), the $\mathcal{E}(\bar{P}_1)$ is:

$$\begin{aligned} &(1-\hat{X}_1(\bar{P}_1))^2 P_1(1-q_1^*) + (\hat{X}_1(\bar{P}_1))^2(1-P_1)q_0^* \\ &+(1-\hat{X}_0(\bar{P}_1))^2(1-P_1)q_0^* + (\hat{X}_0(\bar{P}_1))^2(1-P_1)(1-q_0^*). \end{aligned} \tag{15}$$

If picking $\bar{P}_1 = P_1$, $\mathcal{E}(P_1)$ becomes:

$$\begin{aligned} &(1-\hat{X}_1(P_1))^2 P_1(1-q_1^*) + (\hat{X}_1(P_1))^2(1-P_1)q_0^* \\ &+(1-\hat{X}_0(P_1))^2(1-P_1)q_0^* + (\hat{X}_0(P_1))^2(1-P_1)(1-q_0^*). \end{aligned} \tag{16}$$

The distance of $\mathcal{E}(\bar{P}_1)$ and $\mathcal{E}(P_1)$ is calculated accordingly. □

From Lemma 1, we can learn that the optimal $\bar{P}_1^*$ when $[a,b]$ and ϵ are fixed is P_1, when $\bar{P}_1 \neq P_1$, the MSE increases as a result, and the increased amount is proportional to the square of $|\bar{P}_1 - P_1|$, hence, we need to find the $\bar{P}_1^*$ that cannot be far away from P_1, thus we have the following proposition.

Theorem 5. *By Lemma 1, for a given range $[a,b]$, the optimal $\bar{P}^*$ is $\bar{P}^* = \frac{a+b}{2}$.*

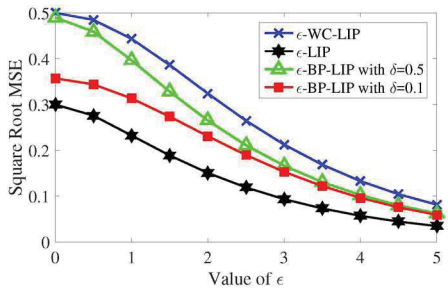

Fig. 4. Numerical analysis on utility-privacy tradeoffs.

Proof. $\mathcal{E}(\bar{P}_1)$ can be represented as:

$$\mathcal{E}(\bar{P}_1) = \mathcal{E}(P_1) + \frac{(e^\epsilon - 1 + b)b(e^\epsilon - 1)^2}{(e^\epsilon - a)(1-a)e^{3\epsilon}}(P_1 - \bar{P}_1)^2, \tag{17}$$

where the first part is not determined by $\bar{P}_1$, thus $\min\max \mathcal{E}(\bar{P}_1)$ is equivalent to $\min\max(\bar{P}_1 - P_1)^2$. The result is straightforward. □

Notice that, in finding the optimal $\bar{P}_1^*$, there are other types of algorithms other than the minimax formulation, such as minimize the expected MSE:

$$\min_{\bar{P}_1^* \in [a,b]} E_{P_1 \in \{a,b\}}[\mathcal{E}(P_1, \bar{P}_1)]. \tag{18}$$

The steps in finding the optimal $\bar{P}_1^*$ are similar, in (18), the problem is equivalent to finding $\bar{P}_1$ that minimize $E[(P_1 - \bar{P}_1)^2]$, and the optimal $\bar{P}_1^* = E[P_1]$. When P_1 is uniformly distributed in $[a,b]$, $\bar{P}_1^* = \frac{a+b}{2}$.

By Theorem 4 and Theorem 5, when the range of P_1: $[a,b]$ is fixed, the optimal $\bar{P}_1^* = \frac{a+b}{2}$ and the optimal perturbation parameters $q_0^* = \frac{b}{b-a+e^\epsilon}$ and $q_1^* = \frac{1-a}{b-a+e^\epsilon}$.

IV. Simulation

In this section, we simulate with numerical results and real world data to show the advantages of the proposed variations of LIP. For the first part, we use numerical results to show the impact of the prior uncertainty to the utility-privacy tradeoffs and compare the performance of different privacy notions. In the second part, we use real-world data to illustrate the advantages of the BP-LIP comparing with LDP.

The first experiment is the comparison among different privacy notions with numerical analysis, including ϵ-WC-LIP, ϵ-LIP, ϵ-BP-LIP. For the ϵ-BP-LIP, we compare two cases with different prior uncertainties, the uncertainty of the prior is measured by $\delta = b - a$. We randomly generate a prior P_1 and generate the user's data from a binary domain according to P_1, then we randomly generate a and b such that $a \leq P_1 \leq b$ under two cases: $\delta = 0.5$ and $\delta = 0.1$. Then we calculate the MSEs which are averaged 5000 times with different a and b. For the WC-LIP, we fix $a = 0$ and $b = 1$. The utility-privacy tradeoffs under different privacy notions are shown in Fig. 4.

Observe from Fig. 4, the ϵ-LIP achieves the best utility while the ϵ-WC-LIP achieves the worst utility under each ϵ. This is because the worst-case LIP has fixed (smaller) feasible region for parameters than LIP. Other than that, the utility privacy tradeoffs of the ϵ-BP-LIP is sandwiched between ϵ-LIP

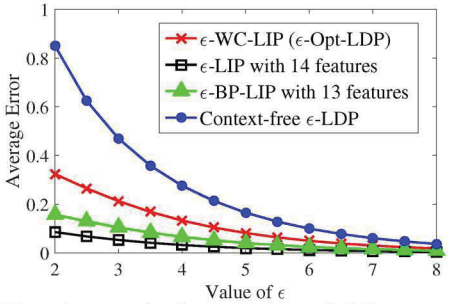

Fig. 5. Utility-privacy tradeoffs comparison of different notions with the income consensus dataset.

and ϵ-WC-LIP, and the curve with larger uncertainty results in a decreased utility.

Next, we testify our analysis by simulation on a real world dataset: "Census income" (Adult dataset). This is a census survey dataset in which 48842 users' personal information are listed, including 14 attributes, such as: age, work class, marriage, race, sex, education and annual income. In the field of machine learning, the Adult dataset is usually used for predicting whether each user's annual income is over 50k dollar by training on all the personal information (taken as features). This is an application of data aggregation for multiple users. As we discussed in Sec. II, when each user is assumed to independently publish data, the optimal parameters for each user are achieved at their local optimality. i.e., if each user perturbs data using the optimal parameters derived in Sec. III, the overall aggregation result is also optimal.

Each user firstly converts his annual income as $\{0, 1\}$, by the following rule that: if income is above 50k, his data is 1 and 0 otherwise. We first create a deep neural network using all the 14 features to predict each user's the annual income. Each prediction probability is used as the exact prior. Then users are perturbing data according to the optimal mechanism of LIP ($(q_0, q_1) = (\frac{P_1}{e^\epsilon}, \frac{1-P_1}{e^\epsilon})$). At last we calculate the average error between the estimation and the raw data. This case is denoted as the ϵ-LIP in Fig. 5. Similarly, we then derive optimal parameters under $a = 0$ and $b = 1$ and denote it as ϵ-WC-LIP. To create a bounded-prior range, we train the network with arbitrary 13 features, and collect prediction priors from all the 14 feature-combinations, thus forming a prior family $\mathbb{P}_1 = \{P_1^1, P_1^2, ..., P_1^{14}\}$. Then $a = \min \mathbb{P}_1$ and $b = \max \mathbb{P}_1$. The following steps are similar, but with BP-LIP optimal parameters. The last notion is context-free LDP, which is with the same mechanism as WC-LIP. There are two cases under the LDP notion, the first one is with prior-related estimator, which is the same with other notions. This case is denoted as the Opt-LIP, which is equivalent to WC-LIP. The second case is with prior-free estimator proposed in [2]: $\hat{X}_{LDP} = (1 - q)Y + \frac{q}{2}$.

From Fig. 5, we can observe similar results as shown in the numerical result. However, in this case, the exact prior trained with 14 features is very accurate, thus ϵ-LIP performs much better than in the numerical result. Other than that, the Opt-LDP leads to an increased utility than the context-free LDP, which means prior-related estimator is a promising approaching to increase utility.

V. Conclusion

In this paper, the notions of worst-case LIP and bounded-prior LIP are proposed to overcome the limitation of LIP and can be implemented according to different scenarios: LIP can be used when the exact prior is available; when there is uncertainty on the prior, BP-LIP can be used instead; If there is no prior knowledge, WC-LIP can be used alternatively. We also derive the relationships between each notion with LDP, and show that ϵ-WC-LIP is equivalent to ϵ-LDP, and ϵ-BP-LIP is sandwiched between ϵ-LDP and 2ϵ-LDP. Then we model a privacy preserving binary data collection mechanism, whose parameters are derived from an optimization problem with the goal to minimize the mean square error subject to the privacy constrains. The optimal solutions are derived in closed form and allow us to characterize the utility-privacy tradeoffs for this problem. At last, we present simulation results for both synthetic data and real world data to illustrate the advantages of the LIP and BP-LIP models.

References

[1] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 838–852, June 2013.

[2] 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.

[3] T. Wang, N. Li, and S. Jha, "Locally differentially private frequent itemset mining," in *2018 IEEE Symposium on Security and Privacy (SP)*, vol. 00, pp. 578–594.

[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] 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.

[6] 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.

[7] 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*, pp. 903–908, Sept 2014.

[8] 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.

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

[10] A. Greenberg, "Apples differential privacy is about collecting your databut not your data," 2016.

[11] 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.

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

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

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

[15] B. Jiang, M. Li, and R. Tandon, "Context-aware data aggregation with localized information privacy," in *2018 IEEE Conference on Communications and Network Security (CNS)*, pp. 1–9, May 2018.

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

APPENDIX A
PROOF OF THEOREM 3

Proof. Suppose a mechanism $\mathcal{M}$ satisfies ϵ-BP-LIP, which means $\forall x, y \in \{0,1\}$, there is:

$$\max_{P_1\in[a,b]} \frac{Pr(Y=y)}{Pr(Y=y|X=1)} \le e^{\epsilon} \tag{19}$$

$$\max_{P_1\in[a,b]} \frac{Pr(Y=y)}{Pr(Y=y|X=0)} \le e^{\epsilon} \tag{20}$$

$$e^{-\epsilon} \le \min_{P_1\in[a,b]} \frac{Pr(Y=y)}{Pr(Y=y|X=1)}, \tag{21}$$

$$e^{-\epsilon} \le \min_{P_1\in[a,b]} \frac{Pr(Y=y)}{Pr(Y=y|X=0)}, \tag{22}$$

Eq. (19) gives us (by Bayes rule):

$$\max_{P_1\in[a,b]} P_1 + \frac{Pr(Y=y|X=0)(1-P_1)}{Pr(Y=y|X=1)} \le e^{\epsilon}.$$

Denote $L_{01} = \frac{Pr(Y=y|X=0)}{Pr(Y=y|X=1)}$.

When $L_{01} > 1$, to find the max, we want $(1-P_1)$ as large as possible, thus $P_1 = a$, $(1-P_1 = 1-a)$. Hence,

$$a + (1-a)L_{01} \le e^{\epsilon}. \tag{23}$$

Thus, $1 < L_{01} \le \frac{e^{\epsilon}-a}{1-a}$.

When $L_{01} \le 1$, to find the max, we want $(1-P_1)$ as small as possible, thus $P_1 = b$, $(1-P_1 = 1-b)$. Hence,

$$b + (1-b)L_{01} \le e^{\epsilon}. \tag{24}$$

Thus, $L_{01} \le \frac{e^{\epsilon}-b}{1-b}$ and $L_{01} \le 1$, the intersection is $L_{01} \le 1$. Which means when $L_{01} \le 1$, all values satisfy the constraints.

Combine the two cases, we have: $L_{01} \le \frac{e^{\epsilon}-a}{1-a}$.

Similarly, constraint (20) gives us:

$$(1-P_1) + P_1 L_{10} \le e^{\epsilon}, \tag{25}$$

As a result, we have: $L_{10} \le \frac{e^{\epsilon}+b-1}{b}$.

Eq. (19) gives us (by Bayes rule):

$$e^{-\epsilon} \le \min_{P_1\in[a,b]} P_1 + \frac{Pr(Y=y|X=0)(1-P_1)}{Pr(Y=y|X=1)}.$$

When $L_{01} > 1$, to find the minimum, we want $(1-P_1)$ as small as possible, thus $P_1 = b$, $(1-P_1 = 1-b)$. Hence,

$$e^{-\epsilon} \le b + (1-b)L_{01} \tag{26}$$

Thus, $L_{01} \ge \frac{e^{-\epsilon}-b}{1-b}$ and $L_{01} > 1$, the intersection is $L_{01} > 1$. Which means when $L_{01} > 1$, all values satisfy the constraints.

When $L_{01} \le 1$, to find the minimum, we want $(1-P_1)$ as large as possible, thus $P_1 = a$, $(1-P_1 = 1-a)$. Hence,

$$e^{-\epsilon} \le a + (1-a)L_{01} \tag{27}$$

Thus, $\frac{e^{-\epsilon}-a}{1-a} \le L_{01} \le 1$.

Combine the two cases, we have: $\frac{e^{-\epsilon}-a}{1-a} \le L_{01}$.

When $e^{-\epsilon} - a \le 0$, the in-equation always hold. When $e^{-\epsilon} - a > 0$, we have $L_{10} = 1/L_{01} \le \frac{1-a}{e^{-\epsilon}-a}$.

Similarly, constraint (22) gives us: When $e^{-\epsilon} - 1 + b > 0$, $L_{01} = 1/L_{10} \le \frac{b}{e^{-\epsilon}-1+b}$.

In summary, there are two cases regarding the relation between $a+b$ and 1, and under each case, there are three cases on the e^{ϵ} and the boundary:

When $a \le 1-b$**:** a is the first boundary when $e^{-\epsilon}$ approaches 1, and $1-b$ is the second.

Case 1 $e^{-\epsilon} \le a \le 1-b$**:** $L_{01} \le \frac{e^{\epsilon}-a}{1-a}$ and $L_{10} \le \frac{e^{\epsilon}+b-1}{b}$. where: $\frac{e^{\epsilon}-a}{1-a} - \frac{e^{\epsilon}+b-1}{b} = \frac{(e^{\epsilon}-1)(a+b-1)}{b(1-a)} \le 0$ So we have:$L_{01}, L_{10} \le \frac{e^{\epsilon}+b-1}{b}$, as $\max\{\frac{e^{\epsilon}-a}{1-a}, \frac{e^{\epsilon}+b-1}{b}\} = \frac{e^{\epsilon}+b-1}{b}$

Case 2 $a \le e^{-\epsilon} \le 1-b$**:** $L_{01} \le \frac{e^{\epsilon}-a}{1-a}$ and $L_{10} \le \min\{\frac{e^{\epsilon}+b-1}{b}, \frac{1-a}{e^{-\epsilon}-a}\}$. To find the minimum of L_{10}: $\frac{e^{\epsilon}+b-1}{b} - \frac{1-a}{e^{-\epsilon}-a} = \frac{(e^{\epsilon}-1)(ae^{\epsilon}-1+b)}{b(e^{-\epsilon}-a)}$.

when $e^{\epsilon} > \frac{1-b}{a}$: $\min\{\frac{e^{\epsilon}+b-1}{b}, \frac{1-a}{e^{-\epsilon}-a}\} = \frac{e^{\epsilon}+b-1}{b}$;

when $e^{\epsilon} \le \frac{1-b}{a}$: $\min\{\frac{e^{\epsilon}+b-1}{b}, \frac{1-a}{e^{-\epsilon}-a}\} = \frac{1-a}{e^{-\epsilon}-a}$;

compare $\frac{1-a}{e^{-\epsilon}-a}$ with $\frac{e^{\epsilon}-a}{1-a}$, we have:

$\frac{e^{\epsilon}-a}{1-a} - \frac{1-a}{e^{-\epsilon}-a} = \frac{a(2-e^{\epsilon}-e^{-\epsilon})}{(1-a)(e^{-\epsilon}-a)} \le 0$. So, in this case:

when $e^{\epsilon} > \frac{1-b}{a}$: $L_{01}, L_{10} \le \frac{e^{\epsilon}+b-1}{b}$;

when $e^{\epsilon} \le \frac{1-b}{a}$: $L_{01}, L_{10} \le \frac{1-a}{e^{-\epsilon}-a}$;

Case 3:$a \le 1-b \le e^{-\epsilon}$**:** $L_{01} \le \min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\}$ and $L_{10} \le \min\{\frac{e^{\epsilon}+b-1}{b}, \frac{1-a}{e^{-\epsilon}-a}\}$. The minimum value of L_{10} is identical to case 2. Comparing $\frac{e^{\epsilon}-a}{1-a}$ with $\frac{b}{e^{-\epsilon}+b-1}$, we have: $\frac{e^{\epsilon}-a}{1-a} - \frac{b}{e^{-\epsilon}+b-1} = \frac{[(b-1)e^{\epsilon}+a](e^{\epsilon}-1)}{(e^{-\epsilon}+b-1)(1-a)} \le \frac{(b-1+a)e^{\epsilon}(e^{\epsilon}-1)}{(e^{-\epsilon}+b-1)(1-a)} \le 0$. Thus $\min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\} = \frac{e^{\epsilon}-a}{1-a}$. And we also showed in last case that $\frac{e^{\epsilon}-a}{1-a} \le \min\{\frac{e^{\epsilon}+b-1}{b}, \frac{1-a}{e^{-\epsilon}-a}\}$. Thus the result of this case is identical to that of case 2:

when $e^{\epsilon} > \frac{1-b}{a}$: $L_{01}, L_{10} \le \frac{e^{\epsilon}+b-1}{b}$;

when $e^{\epsilon} \le \frac{1-b}{a}$: $L_{01}, L_{10} \le \frac{1-a}{e^{-\epsilon}-a}$;

When $a > 1-b$**:** $1-b$ is the first boundary when $e^{-\epsilon}$ approaches 1, and a is the second.

Case 4 $e^{-\epsilon} \le 1-b \le a$**:** $L_{01} \le \frac{e^{\epsilon}-a}{1-a}$ and $L_{10} \le \frac{e^{\epsilon}+b-1}{b}$. where:$\frac{e^{\epsilon}-a}{1-a} - \frac{e^{\epsilon}+b-1}{b}$. $= \frac{(e^{\epsilon}-1)(a+b-1)}{b(1-a)} > 0$ So we have:$L_{01}, L_{10} \le \frac{e^{\epsilon}-a}{1-a}$, as $\max\{\frac{e^{\epsilon}-a}{1-a}, \frac{e^{\epsilon}+b-1}{b}\} = \frac{e^{\epsilon}-a}{1-a}$

Case 5 $1-b \le e^{-\epsilon} \le a$**:** $L_{01} \le \min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\}$ and $L_{10} \le \frac{e^{\epsilon}+b-1}{b}$. To find the minimum of L_{01}: $\frac{e^{\epsilon}-a}{1-a} - \frac{b}{e^{-\epsilon}+b-1} = \frac{(e^{\epsilon}-1)[a-(1-b)e^{\epsilon}]}{(1-a)(e^{-\epsilon}+b-1)}$.

when $e^{\epsilon} > \frac{a}{1-b}$: $\min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\} = \frac{e^{\epsilon}-a}{1-a}$;

when $e^{\epsilon} \le \frac{a}{1-b}$: $\min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\} = \frac{b}{e^{-\epsilon}+b-1}$;

compare $\frac{b}{e^{-\epsilon}+b-1}$ with $\frac{e^{\epsilon}+b-1}{b}$, we have:$\frac{b}{e^{-\epsilon}+b-1} - \frac{e^{\epsilon}+b-1}{b}$ $= \frac{(1-b)(e^{\epsilon}+e^{-\epsilon}-2)}{b(e^{-\epsilon}+b-1)} \ge 0$. So, in this case:

when $e^{\epsilon} > \frac{a}{1-b}$: $L_{01}, L_{10} \le \frac{e^{\epsilon}-a}{1-a}$;

when $e^{\epsilon} \le \frac{a}{1-b}$: $L_{01}, L_{10} \le \frac{b}{e^{-\epsilon}+b-1}$;

Case 6:$1-b \le a \le e^{-\epsilon}$**:** $L_{01} \le \min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\}$ and $L_{10} \le \min\{\frac{e^{\epsilon}+b-1}{b}, \frac{1-a}{e^{-\epsilon}-a}\}$. The minimum value of L_{01} is identical to case 5.

Comparing $\frac{e^{\epsilon}+b-1}{b}$ with $\frac{1-a}{e^{-\epsilon}-a}$, we have: $\frac{e^{\epsilon}+b-1}{b} - \frac{1-a}{e^{-\epsilon}-a} = \frac{(e^{\epsilon}-1)(-ae^{\epsilon}-b+1)}{(e^{-\epsilon}-a)b} \le \frac{(e^{\epsilon}-1)(1-b-a)e^{\epsilon}}{(e^{-\epsilon}-a)b} \le 0$ Thus $\min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\} = \frac{e^{\epsilon}+b-1}{b}$. And we also showed in last case that $\frac{e^{\epsilon}+b-1}{b} \le \min\{\frac{e^{\epsilon}-a}{1-a}, \frac{b}{e^{-\epsilon}+b-1}\}$. Thus the result of this case is identical to that of case 5:

when $e^{\epsilon} > \frac{a}{1-b}$: $L_{01}, L_{10} \le \frac{e^{\epsilon}-a}{1-a}$;

when $e^{\epsilon} \le \frac{a}{1-b}$: $L_{01}, L_{10} \le \frac{b}{e^{-\epsilon}+b-1}$;

In summary, the above cases can combine into four cases as shown in the theorem. □