Local Differential Privacy for Belief Functions

Qiyu Li, Chunlai Zhou, Biao Qin and Zhiqiang Xu

1Computer Science Dept., Renmin University of China, Beijing, CHINA
2Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, UAE
{qiyuli, czhou, qinbiao}@ruc.edu.cn, zhiqiangxu2001@gmail.com

Abstract
In this paper, we propose two new definitions of local differential privacy for belief functions. One is based on Shafer’s semantics of randomly coded messages and the other from the perspective of imprecise probabilities. We show that such basic properties as composition and post-processing also hold for our new definitions. Moreover, we provide a hypothesis testing framework for these definitions and study the effect of “don’t know” in the trade-off between privacy and utility in discrete distribution estimation.

1 Introduction
Differential privacy (DP) is a mathematically rigorous definition of privacy which addresses the paradox of learning nothing about an individual while learning useful information about a population [12, 13]. In particular, local differential privacy (LDP) is a model of differential privacy with the added restriction that even if an adversary has access to the personal responses of an individual in the database, that adversary will still be unable to learn too much about the user’s personal data [22, 20, 10]. The uncertainty in standard LDP mechanisms is usually provided by randomization which associates each input with a probability function over all possible outputs. The prototypical example of an LDP mechanism is the randomized response survey technique proposed in [35]. Current randomized response mechanisms equate privacy-preserving with lying and are designed on the assumption that users abide by the data collection protocol which allows respondents to lie with a known probability. However, recent research results from the perspective of the respondents show that, in practice, although these mechanisms allow the respondents to maintain privacy, the procedures may confuse respondents, fail to address the concerns of the users and hence yield nonresponse or noncompliance [37, 4, 29]. An effective differential privacy communication can increase data-sharing rates [37].

To address noncompliance and nonresponse, we propose in this paper to design differential privacy mechanisms which incorporate “don’t know” or nonresponse as an alternative outcome or allow imprecision in the mechanism design. In practice, people may prefer not to respond or say “I don’t know” to withhold sensitive information which minimizes the questionable ethical consequences of
lying in their eyes [3]. By addressing such ethical privacy concerns, our new mechanisms aims to increase respondents’ willing to share their data. Here we study this new type of privacy mechanisms from a more general Dempster-Shafer perspective by representing uncertainty in privacy mechanisms with belief functions [6][30]. The Dempster-Shafer theory (also known as the theory of evidence or the theory of belief functions) is a well-known uncertainty theory for its expressiveness in representing ignorance. The theory improves the root concepts of probabilities “yes” and “no” that sum to one, by appending a third probability of “don’t know” [7]. As the world of statistical analysis moves more and more to “big data” and associated “complex systems”, the Dempster-Shafer theory provides a middle ground with the third probability “don’t know” and can be expected to become increasingly important in privacy protection.

Our first and main contribution in this paper is to propose two new definitions of LDP (one is $\epsilon$-local differential privacy according to Shafer ($\epsilon$-SLDP) (Definition 3.1) and the other according to Walley ($\epsilon$-WLDP) (Definition 3.13)) and to provide a statistical framework for these two definitions as the trade-offs between type I and II errors in a natural hypothesis-testing problem (Theorems 3.5 and 3.18). Our second contribution is to characterize the effect of “don’t know” in the trade-off between privacy and utility in discrete distribution estimation problem. The privacy mechanisms in the two definitions associate each input $x$ with a belief function on the output set $\hat{Y}$. The difference between these two definitions comes from their different semantics of belief functions. The first definition is motivated by Shafer’s interpretation of belief functions as randomly coded messages [32]. In this semantics, we generalize Warner’s randomized response mechanism by allowing answering “don’t know” with probability $1 - p - q$ where $p$ is the probability of answering truthfully and $q$ the probability of lying. For the discrete distribution estimation problem of a generalized Warner’s model, we study the effect of “don’t know” on the trade-off between the privacy loss and the estimation accuracy. The most important and difficult step is to compute the variance of the maximum likelihood estimation of the parameter $\pi$, the true proportion of the people with the sensitive property. We employ some combinatorial techniques to obtain a formula for the estimation accuracy (Theorem 3.10). We show that, when the probability of “don’t know” increases, the overall effect of the trade-off for this generalized model decreases, and when this probability equals 0, the effect is optimal and the trade-off is the same as that for the standard Warner’s model (Figure 2). In the second definition, we adopt the imprecise-probability semantics to accommodate unknown response probabilities in privacy mechanisms and interpret belief function $bel$ as the set of all probability functions $pr$ which are consistent with $bel$ [33]. Both the privacy loss and estimation accuracy are defined with respect to those consistent probability functions according to the worst-case analysis. Moreover, we compare the trade-offs between privacy and estimation accuracy for these two definitions ($\epsilon$-SLDP and $\epsilon$-WLDP) and Warner’s randomized response mechanism (Figure 5).

2 Dempster-Shafer Theory

Let $\Omega$ be a frame and $\mathcal{A} = 2^\Omega$ be the Boolean algebra of propositions. $|A|$ denotes the cardinality of a subset $A$. A mass assignment (or mass function) over $\Omega$ is a mapping $m : \mathcal{A} \rightarrow [0, 1]$ satisfying $\sum_{A \in \mathcal{A}} m(A) = 1$. A mass function $m$ is called normal if $m(\emptyset) = 0$. A belief function is a function $bel : \mathcal{A} \rightarrow [0, 1]$ satisfying the conditions: $bel(\emptyset) = 0$, $bel(\Omega) = 1$ and $bel(\bigcup_{i=1}^n A_i) \geq \sum_{\emptyset \neq I \subseteq \{1, \ldots, n\}} (-1)^{|I|+1}bel(\bigcap_{i \in I} A_i)$ where $A_i \in \mathcal{A}$ for all $i \in \{1, \ldots, n\}$. A mapping $f : \mathcal{A} \rightarrow [0, 1]$
is a belief function if and only if its Möbius transform is a mass assignment (Page 39 in [30]). In other words, if \( m : \mathcal{A} \to [0, 1] \) is a mass assignment, then it determines a belief function \( \text{bel} : \mathcal{A} \to [0, 1] \) as follows: \( \text{bel}(A) = \sum_{B \subseteq A} m(B) \) for all \( A \in \mathcal{A} \). Moreover, given a belief function \( \text{bel} \), we can obtain its corresponding mass function \( m \) as follows: \( m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \text{bel}(B) \) for all \( A \in \mathcal{A} \). Intuitively, for a subset event \( A \), \( m(A) \) measures the belief that an agent commits exactly to \( A \), not the total belief \( \text{bel}(A) \) that an agent commits to \( A \). A subset \( A \) with non-zero mass is called a focal set. The belief function \( \text{bel} \) is called Bayesian if \( m(A) = 0 \) for all non-singletons \( A \). The corresponding plausibility function \( \text{pl}_m : 2^\Omega \to [0, 1] \) is defined by \( \text{pl}_m(A) = \sum_{E \cap A \neq \emptyset} m(E) \) for all \( A \subseteq \Omega \). Whenever the context is clear, we drop the subscript \( m \). For \( m, \text{bel} \) and \( \text{pl}_m \), if we know any one of them, then we can determine the other two. Without further notice, all mass functions in this paper are assumed to be normal and all subsets are focal.

In this paper, we focus on only two semantics of belief functions. The first one is Shafer’s semantics of belief functions in terms of randomly coded messages. Suppose someone chooses a code at random from a list of codes, uses the code to encode a message, and then sends us the result. We know the list of codes and the chance of each code being chosen—say the list is \( c_1, \ldots, c_n \), and the chance of \( c_i \) being chosen is \( p_i \). We decode the encoded message using each of the codes and find that this always produces a message of the form “the truth is in \( A \)” for some non-empty subset \( A \) of the set of possibilities \( \Omega \). Let \( A_i \) denote the subset we get when we decode using \( c_i \), and set \( m(A) = \sum \{ p_i : 1 \leq i \leq n, A_i = A \} \) for each \( A \subseteq \Omega \). The number \( m(A) \) is the sum of the chances for those codes that indicate \( A \) was the true message; it is, in a sense, the total chance that the true message implies \( A \). If the true message is infallible and the coded message is our only evidence, then it is natural to call \( \text{bel}(A) \) our probability or degree of belief that the truth lies in \( A \). The second interpretation of belief functions in this paper is from the perspective of imprecise probabilities. Given a belief function \( \text{bel} \), let \( \mathcal{P}_{\text{bel}} \) denote the set of all probability functions which are consistent with or dominate over \( \text{bel} \). In other words, \( \mathcal{P}_{\text{bel}} = \{ pr : pr \text{ is a probability function on } \Omega \text{ and } pr \geq \text{bel} \} \) where \( pr \geq \text{bel} \) means \( pr(E) \geq \text{bel}(E) \) for all \( E \subseteq \Omega \). Due to lack of information, uncertainty can’t be represented by a probability function but by a belief function \( \text{bel} \). All consistent probability functions are possible. Whenever enough information is available, we may specify a probability function from \( \mathcal{P}_{\text{bel}} \) to represent the uncertainty. One may refer to [5] and [13] for a detailed introduction to belief functions and DP.

### 3 Local Differential Privacy

Let \( X \) be a private source of information defined on a discrete, finite input alphabet \( X = \{ x_1, \ldots, x_k \} \) and \( Y \) be an output alphabet \( Y = \{ y_1, \ldots, y_l \} \) that need not be identical to the input alphabet \( X \). In this paper, we will represent a privacy mechanism \( Q \) via a row-stochastic matrix. For simplicity, we also use \( Q \) to denote this matrix. \( Q \) is called an evidential privacy mechanism if each row of the matrix \( Q \) is a mass function on \( Y \). In other words, each evidential privacy mechanism \( Q \) maps \( X = x \) to \( Y \in E \) with \( Q(x) \) which can be represented by a mass \( \text{bel}^Q_x(E) \) (belief \( \text{bel}^Q_x(E) \) or plausibility \( \text{pl}^Q_x(E) \)) where \( \text{bel}^Q_x(E) \) (belief \( \text{pl}^Q_x(E) \) or plausibility \( \text{pl}^Q_x(E) \)) is a mass (belief or plausibility) function on \( Y \) for all \( x \in X \). Since \( \text{bel}^Q_x(\emptyset) = 0 \) for all \( x \), we write the mechanism \( Q \) as a \( k \times (2^l - 1) \) matrix. Whenever the context is clear, we usually drop the superscript \( Q \). In this paper, we assume that all the alphabet sets are finite.
In other words, an evidential privacy mechanism is just a standard LDP mechanism whose instructions are defined by random sets instead of probability functions.

### 3.1 LDP according to Shafer

For an evidential privacy mechanism $Q$, let $r^Q_S = \max_{x, x' \in X \subseteq Y} \frac{m^Q_x(E)}{m^Q_{x'}(E)}$ and $\epsilon^Q_S = \ln(r^Q_S)$.

**Definition 3.1** For any $\epsilon > 0$, the mechanism $Q$ is called $\epsilon$-locally differential private according to Shafer ($\epsilon$-SLDP for short) if $-\epsilon \leq \epsilon^Q_S \leq \epsilon$. And $\epsilon^Q_S$ is called the privacy loss of $Q$ according to Shafer and $\epsilon$ is a privacy budget.

In other words, by observing $E$, the adversary cannot reliably infer whether $X = x$ or $X = x'$ (for any pair $x$ and $x'$). Indeed, the smaller the $\epsilon$ is, the closer the likelihood ratio of $X = x$ to $X = x'$ is to 1. Therefore, when $\epsilon$ is small, the adversary cannot recover the true value of $X$ reliably. In this definition, we adopt Shafer’s interpretation as randomly coded messages. Each subset of $Y$ is treated as an individual message or response. The mechanism randomly chooses a code $c$ and uses it to encode a message $E \subseteq Y$. And $m_x(E)$ is equal to the chance of choosing $c$. If we set $2^Y \setminus \{\emptyset\}$ as the output alphabet, then the above $Q$ is simply the standard local differential private mechanism. In particular, if each row of $Q$ is Bayesian, then $Q$ is essentially a standard randomized mechanism and the $\epsilon$-SLDP is just the standard $\epsilon$-LDP for randomized privacy mechanisms. Almost all basic properties for privacy-preserving randomized mechanisms can be generalized to the setting of belief functions.

Let $r_{pl,S}^Q = \max_{x, x' \in X \subseteq Y} \frac{pl^Q_x(E)}{pl^Q_{x'}(E)}$ and $r_{bel,S}^Q = \max_{x, x' \in X \subseteq Y} \frac{bel^Q_x(E)}{bel^Q_{x'}(E)}$. Denote $\epsilon_{pl,S}^Q := \ln(r_{pl,S}^Q)$ and $\epsilon_{bel,S}^Q := \ln(r_{bel,S}^Q)$.

**Lemma 3.2** If privacy mechanism $Q$ is $\epsilon$-SLDP, then $-\epsilon \leq \epsilon_{bel,S}^Q \leq \epsilon$ and $-\epsilon \leq \epsilon_{pl,S}^Q \leq \epsilon$.

From Lemma 3.2 we know that $\epsilon^Q_S \geq \epsilon_{pl,S}^Q$. But generally we don’t have the converse that $\epsilon_{pl,S}^Q \geq \epsilon^Q_S$. If we have several building blocks for designing differentially private algorithms, it is important to understand how we can combine them to design more sophisticated algorithms.

**Lemma 3.3** (Composition) Let $Q_1$ be an $\epsilon_1$-SLDP evidential privacy mechanism from $X$ to $Y_1$ and $Q_2$ be an $\epsilon_2$-SLDP evidential privacy mechanism from $X$ to $Y_2$. Then their combination $Q_{1,2}$ defined by $Q_{1,2}(x) = (Q_1(x), Q_2(x))$ is $\epsilon_1 + \epsilon_2$-SLDP.

The composition of a data-independent mapping $f$ with an $\epsilon$ locally differential private algorithm $Q$ is also $\epsilon$ locally differential private.

**Lemma 3.4** (Post-processing) Let $Q$ be an $\epsilon$-SLDP mechanism from $X$ to $Y$ and $f$ is a randomized algorithm from $Y$ to another finite alphabet set $Z$. Then $f \circ Q$ is an $\epsilon$-SLDP mechanism from $X$ to $Z$.

Now we offer a hypothesis testing interpretation for the above $\epsilon$-SLDP. From an attacker’s perspective, the privacy requirement can be formalized as the following hypothesis testing problem for two datasets $x$ and $x'$:

$H_0$: the underlying dataset is $x$ vs. $H_1$: the underlying dataset is $x'$.
The output of the mechanism \( Q \) serves as the basis for performing the hypothesis testing problem. The distinguishability of the two inputs \( x \) and \( x' \) can be translated into the trade-off between type I and type II errors [9]. For belief functions, it is natural to consider minimax tests [18]. Formally, consider a rejection rule \( \varphi : Y \rightarrow [0, 1] \). Let \( \mathcal{P}^Q_x \) and \( \mathcal{P}^Q_{x'} \) denote the two sets of probability functions dominating \( \text{bel}^Q_x \) and \( \text{bel}^Q_{x'} \) respectively. In other words, \( \mathcal{P}^Q_x = \{ p_r \in \Delta(Y) : p_r \geq \text{bel}^Q_x \} \) and \( \mathcal{P}^Q_{x'} = \{ p_r \in \Delta(Y) : p_r \geq \text{bel}^Q_{x'} \} \). The lower power of \( \varphi \) under \( x' \) is defined as \( \pi_{x'} := \inf_{p_r \in \mathcal{P}^Q_{x'}} E_p(p_r(\varphi)) \).

In the setting of \( \varepsilon\text{-SLDP} \), we assume that type I error \( \alpha_\varphi \) is represented by \( \sup_{p_r \in \mathcal{P}^Q_x} E_p(p_r(\varphi)) \) and type II error by \( \beta_\varphi = 1 - \inf_{p_r \in \mathcal{P}^Q_{x'}} E_p(p_r(\varphi)) \). A test \( \varphi \) is called a level-\( \alpha \) minimax test if \( \varphi = \arg\min \{ \beta_\varphi : \alpha_\varphi \leq \alpha \} \).

The following theorem is a generalization of the well-known result (Theorem 2.4 in [36]) for standard differential privacy.

**Theorem 3.5** For any evidential privacy mechanism \( Q \), the following two statements are equivalent:

1. \( Q \) is \( \varepsilon\text{-SLDP} \);

2. If type I error \( \alpha_\varphi \in [l, L] \), then type II error \( \beta_\varphi \in [u(L), U(l)] \) where \( u(\alpha) := \max \{ e^{-\varepsilon}(1 - \alpha), 1 - \alpha e^{-\varepsilon} \} \) and \( U(\alpha) := \min \{ e^{\varepsilon}(1 - \alpha), 1 - \alpha e^{-\varepsilon} \} \).

Now we consider the hypothesis testing problem for the composition and would like to distinguish between \( Q(x) \times Q(x) \) and \( Q(x') \times Q(x') \). The corresponding type I and II errors \( \alpha_\varphi^2 \) and \( \beta_\varphi^2 \) can be defined similarly. For simplicity, we only show the two-fold composition and other multi-fold compositions can be obtained similarly.

**Corollary 3.6** For the hypothesis testing problem for the composition, if type I error \( \alpha_\varphi^2 \in [l, L] \), then type II error \( \beta_\varphi^2 \in [u^2(L), U^2(l)] \) where \( u^2(\alpha) := \max \{ e^{-2\varepsilon}(1 - \alpha), -\alpha + \frac{2}{e^\varepsilon + 1}, 1 - \alpha e^{-2\varepsilon} \} \) and \( U^2(\alpha) := \min \{ e^{2\varepsilon}(1 - \alpha), 1 - \alpha e^{-2\varepsilon}, -\alpha + \frac{3 - e^{2\varepsilon}}{e^\varepsilon + 1} \} \).

Both Theorem 3.5 and Corollary 3.6 can be visualized in Figure 1.

![Figure 1: Trade-off between type I and II errors for SLDP](image-url)
The discrete estimation problem is defined as follows. Given a prior which is a vector $\pi = (\pi_1, \ldots, \pi_k)$ on the probability simplex $S_k = \{p = (\pi_1, \ldots, \pi_k) : \pi_i \geq 0 (1 \leq i \leq k), \sum_{i=1}^{k} \pi_i = 1\}$, samples $X_1, \cdots, X_n$ are drawn i.i.d. according to $\pi$. A privacy mechanism $Q$ is then applied independently to each sample $X_i$ to produce $Y_n = (Y_1, \cdots, Y_n)$, the sequence of private observations. Observe that the $Y_i$’s are distributed according to $m = \pi Q$, which are mass functions not necessarily probability functions when $Q$ is evidential. Our goal is to estimate the distribution vector $\pi$ from $Y_n$ within a certain privacy budget requirement. The performance of the estimation may be measured via a loss function. Here we use the mean square loss function. $Q$ is called optimal if the estimation error is the smallest. A classic example for discrete distribution estimation is Warner’s randomized response method [35].

Example 3.7 According to prototypical Warner’s randomized response mechanism $Q_W$, the respondent answers truthfully with probability $p$ and lies with probability $1 - p$. Let $\pi$ be the true proportion of the people having property $P$. A sample of $Y_1, \cdots, Y_n$ of respondents are drawn with replacement from the population and their responses are distributed i.i.d. according to $(q_1, q_2) = (\pi, 1 - \pi)Q_W$. So $q_1 = \pi p + (1 - \pi)(1 - p)$ and $q_2 = \pi(1 - p) + (1 - \pi)p$. Arrange the indexing of the sample so that the first $n_1$ respondents say "Yes" and the remaining $n - n_1$ answers "No". We obtain the maximum likelihood estimation of $\pi$ as $\hat{\pi} = \frac{p - 1}{2p - 1} + \frac{n_1}{(p - 1)n}$. It can be shown [35, 16] that this distribution estimation $\hat{\pi}$ is unbiased and its mean square error or variance is the following formula:

$$Var[\hat{\pi}] = \frac{-(\pi - \frac{1}{2})^2 + \frac{1}{4}}{n} + \frac{1}{4(2p - 1)^2} - \frac{1}{4}$$  \hspace{1cm} (1)

Within the privacy budget of $\epsilon$, the optimal privacy mechanism is

$$Q_{WRR} = \frac{1}{e^\epsilon + 1} \begin{pmatrix} e^\epsilon & 1 \\ 1 & e^\epsilon \end{pmatrix}.$$  

Now we are generalizing the above Warner’s model by allowing a third response “I don’t know” and representing the corresponding uncertainty with a mass function. Let $Q_{2 \times 3}$ denote a known row-stochastic matrix as follows:

$$Q_{2 \times 3} = \begin{pmatrix} p & q & 1 - p - q \\ q & 1 - p - q & p \end{pmatrix}$$

where $p, q \in [0, 1]$. $Q_{2 \times 3}$ may be regarded as a generalized Warner’s randomized response mechanism where a respondent answers truthfully with probability $p$, tells a lie with $q$ and don’t respond or respond "I don’t know" with probability $1 - p - q$. We may assume in this paper that $p > \frac{1}{2}$.

Remark 3.8 In the following we choose to work with such a simple form $Q_{2 \times 3}$ of LDP for belief functions. A more general form can be studied similarly, but unfortunately we couldn’t obtain closed forms for (approximate) estimation and error as we achieve below for this simple form $Q_{2 \times 3}$. The maximum likelihood estimation problem for the more general form can be naturally formalized as a mixture of the conditional mass functions associated with the evidential privacy mechanism with the mixture proportions as the unknown prior distribution of the sensitive population. We can apply EM algorithm to approximate the prior distribution and compute its Fisher information and further the
standard error of the approximation \cite{1}. However, the simple form provides us with a neat formula of estimation error (Theorem 3.10) and hence a formula for the privacy-utility trade-off. Indeed the simple form for evidential mechanism is enough to illustrate the effect of the answer “I don’t know” or nonresponse on the privacy-utility trade-off. Both the simulation experiments and Figure 2 afterwards are based on the above analysis. In this paper we mainly focus on this simple form \( Q_{2\times 3} \). But we expect that such a simple form to evidential privacy mechanisms is the same as Warner’s \( 2 \times 2 \) mechanism to the standard LDP. For standard LDP, every approximate DP algorithm can be simulated by a (leaky) variant of Warner’s mechanism to the standard LDP. For standard LDP, every approximate evidential privacy mechanism can be simulated by some variant of our \( 2 \times 3 \) mechanisms in this paper. In this sense, our contribution is similar to Warner’s contribution to standard LDP.

A simple random sample of \( n \) people is drawn with replacement from the population. Let \( Z_i \) denote the \( i \)-th sample element. Recall that \( \pi \) is the true proportion of the people with the sensitive property \( P \). \( Z_i \) is distributed according to the following \((q_1, q_2, q_3)\):

\[
\begin{pmatrix}
q_1 \\
q_2 \\
q_3
\end{pmatrix} = \begin{pmatrix}
\pi & 1 - \pi \\
q & p \\
q & p
\end{pmatrix} \begin{pmatrix}
1 - p - q \\
p \\
1 - p - q
\end{pmatrix}
\]

In other words, \( q_1 = \pi p + (1 - \pi) q \), \( q_2 = \pi q + (1 - \pi) p \), and \( q_3 = 1 - p - q \). Note that \( q_1 + q_2 + q_3 = 1 \). It implies that \( Z_i \) says “Yes”, “No” and “don’t know” with probabilities \( q_1, q_2 \) and \( q_3 \) respectively. Arrange the indexing of the sample so that the first \( n_1 \) sample elements say \( \text{Yes} \), the next \( n_2 \) say \( \text{No} \) and the last \( n_3 \) say “don’t know” where \( n_1, n_2 \) and \( n_3 \) are natural numbers such that \( n_1 + n_2 + n_3 = n \). So the likelihood of the sample is \( L(\pi) = q_1^{n_1} q_2^{n_2} q_3^{n_3} \). By taking its logarithm and then setting its derivative to be zero, we obtain \( \frac{n_1}{q_1} - \frac{n_2}{q_2} = 0 \). So we obtain the maximum likelihood estimation (MLE) of \( \pi \) as follows:

\[
\hat{\pi} = \frac{n_2 q - n_1 p}{(n_1 + n_2)(q - p)}.
\]

Now we want to compute the expectation of \( \hat{\pi} \). From \( Z_i \), we define three new random variables \( Z_{i1} = \mathbb{I}[Z_i = \text{Yes}], Z_{i2} = \mathbb{I}[Z_i = \text{No}] \) and \( Z_{i3} = \mathbb{I}[Z_i = \text{don’t know}] \) (where \( \mathbb{I} \) denotes the indicator function). Then \( Z_i = Z_{i1} + Z_{i2} + Z_{i3}, N_1 = \sum_{i=1}^{n} Z_{i1}, N_2 = \sum_{i=1}^{n} Z_{i2} \) and \( N_3 = \sum_{i=1}^{n} Z_{i3} \). So \( N_1 + N_2 + N_3 = n \). We obtain the conditional expectation of the MLE.

\textbf{Theorem 3.9} \( \mathbb{E}[\hat{\pi}|N_1 + N_2 \neq 0] = \pi \).

\textbf{Theorem 3.10} \( \text{Var}(\hat{\pi}|N_1 + N_2 \neq 0) = \frac{1}{(q - p)^2} \pi p + (1 - \pi)q \pi q + (1 - \pi) p \]

\( A = \left[-(\pi - \frac{1}{2})^2 + \frac{1}{4} \frac{(p - q)^2}{p - q} \right] \) where \( A = \sum_{0 \leq N_3 < n} \frac{n}{n-N_3} \left( \frac{n}{N_3} \right) (1 - q_3)^{n-N_3} q_3^{N_3} \).

The formula in Theorem 3.10 is essential to our analysis of the trade-off between privacy loss and estimation accuracy. One may refer to the supplementary materials for a detailed proof (of independent interest). In this paper, we adopt from \cite{14} a good approximation of \( A \) as \( \frac{1}{(n+1)(p+q)-1} \). In particular, with this approximation, when \( p + q = 1 \), \( \text{Var}(\hat{\pi}|N_1 + N_2 \neq 0) = \frac{-(\pi - \frac{1}{2})^2 + \frac{1}{4} \frac{1}{(p-1)^2}}{n} \), which is exactly the estimation error of Warner’s model (Eq. 1).
Corollary 3.11  Let \( f(q) = \frac{\frac{1}{2} - \frac{1}{2} \left( \frac{p}{q} \right)^2}{n+1} \). Then \( f'(q) > 0 \). In other words, \( \text{Var}(\hat{\pi}) \) is increasing with respect to \( q \).

This proposition tells us that, within the privacy budget of \( \epsilon \), one can increase the estimation accuracy by saying “I don’t know” as much as possible instead of lying.

Corollary 3.12  Fix \( p + q = c \). The optimal \( \epsilon \)-LDP mechanism is

\[
Q_{GWRR} = \left( \begin{array}{c|c|c} \frac{e^c}{e^c + 1} & \frac{c}{e^c + 1} & 1 - c \\ \frac{e^c}{e^c + 1} & \frac{c}{e^c + 1} & 1 - c \end{array} \right)
\]

In order to emphasize the dependency of the privacy matrix \( Q_{2 \times 3} \) on the parameters \( p \) and \( q \), we denote \( Q_{2 \times 3} \) as \( Q_{2 \times 3}(p, q) \), the privacy loss \( \ln\left(\frac{p}{q}\right) \) as \( \epsilon^S(p, q) \) and the estimation error \( \text{Var}(\hat{\pi}|N_1 + N_2 \neq 0) \) as \( \nu^S(p, q) \).

This trade-off formula can be actually easily obtained. What we can achieve is an analysis rather than simulation. Let \( p + q = c \) and \( e^c = \frac{p}{q} = \frac{c}{1 - c - p} \). So we get \( p = \frac{1 - c}{e^c + 1} \). If we substitute this formula into the error formula in Theorem 3.10, then we get a formula of estimation error in terms of the privacy loss. Simulation experiments are carried out to verify the trade-off in the privacy mechanism. In order to reduce the sampling error on the experimental results, the following results are the average of 1000 experimental outcomes.

Figure 2: The trade-off in Shafer’s semantics

The trade-off between the privacy loss \( \epsilon^S(p, q) \) and the accuracy \( \nu^S(p, q) \) can be illustrated in the following Figure 1. The figure shows clearly the impact of “don’t know” with probability \( 1 - p - q \) on the trade-off between \( \epsilon^S(p, q) \) and \( \nu^S(p, q) \). When \( 1 - p - q = 0 \) or \( p + q = 1 \), the black curve for the trade-off between \( \epsilon^S(p, q) \) and \( \nu^S(p, q) \) is exactly for Warner’s randomized response mechanism. If \( p + q = c \) where \( c \) is a constant, the trade-off curve is similar to that for Warner’s mechanism. Moreover, when the constant \( c \) gets smaller or the probability of “don’t know” gets larger, the curve moves further away from that for Warner’s model. Figure 2 tells us that Warner’s model is optimal among those generalized \( Q_{2 \times 3} \)-mechanisms. Next we explore the effect of the sample size on the accuracy of the estimation. We set the sample size to be 10, 100, 500, 1000 and fix \( q_3 = 0.1 \). From the experimental results (Figure 3), we can see that when the privacy loss is relatively large, different
sample sizes can achieve similar estimations. However, when the privacy budget is relatively small, with the increase of the sample size, the estimation variance gets smaller and smaller.

![Figure 3: Impact of sample sizes on the estimation accuracy](image)

3.2 LDP according to Walley

For an evidential privacy mechanism $Q$, let $r^W_Q = \max_{pr_x \in P} \frac{pr_x(E)}{\text{bel}_x(E)}$. And the logarithm $\epsilon^W_Q = \ln(r^W_Q)$ quantifies the privacy loss of the privacy mechanism $Q$ in Walley’s semantics of imprecise probabilities. There is another definition of LDP for belief functions in the setting of imprecise probabilities:

**Definition 3.13** For any $\epsilon > 0$, $Q$ is called $\epsilon$-locally differential private according to Walley ($\epsilon$-WLDP for short) if, $-\epsilon \leq \epsilon^W_Q \leq \epsilon$. And $\epsilon^W_Q$ is called the privacy loss of $Q$ according to Walley and $\epsilon$ is a privacy budget.

In other words, the privacy loss for $\epsilon$-WLDP is defined by consistent probability functions in the worst case. So, $\epsilon$-WLDP fits well with the worst-case analysis behind the philosophy of differential privacy and also with the conservative principle of least commitment in the theory of belief functions [8]. Lemma 3.2 and the following Lemma 3.14 provide a simple mathematical characterization of SLDP and WLDP, where we can see clearly the main difference between Definitions 3.1 and 3.13.

**Lemma 3.14** (Alternative formulations) If privacy mechanism $Q$ is $\epsilon$-WLDP, then, for all $x, x' \in X$ and $E \subseteq Y$: $e^{-\epsilon} \leq \frac{\text{pl}_x(E)}{\text{bel}_x(E)} \leq e^\epsilon$.

**Lemma 3.15** (Composition) Let $Q_1$ be an $\epsilon_1$-WLDP evidential privacy mechanism from $X$ to $Y_1$ and $Q_2$ be an $\epsilon_2$-WLDP evidential privacy mechanisms from $X$ to $Y_2$. Then their combination $Q_{1,2}$ defined by $Q_{1,2}(x) = (Q_1(x), Q_2(x))$ is $\epsilon_1 + \epsilon_2$-WLDP.

**Lemma 3.16** (Post-processing) Let $Q$ be an $\epsilon$-WLDP mechanism from $X$ to $Y$ and $f$ is a data-independent randomized algorithm from $Y$ to another finite alphabet set $Z$. Then $f \circ Q$ is an $\epsilon$-WLDP mechanism from $X$ to $Z$. 
For the hypothesis testing problem, recall that \( Q \) denotes an evidential privacy mechanism and \( \varphi : Y \rightarrow [0, 1] \) is a rejection rule. In order to translate \( \epsilon \)-WLDP into the trade-off between type I and II errors, we have to divide them into two different types of errors: one is pessimistic and the other optimistic. For the rejection rule \( \varphi \), the pessimistic type I and II are defined as \( \alpha^\text{pe}_\varphi = \sup_{\varphi \in P} \mathbb{E}_{\varphi}(\varphi) \) and \( \beta^\text{pe}_\varphi = \sup_{\varphi \in P} \mathbb{E}_{\varphi}(1 - \varphi) \), respectively. They are actually the same as those errors in \( \epsilon \)-SLDP. Also we define the optimistic type I and II errors as \( \alpha^\text{op}_\varphi := \inf_{\varphi \in P} \mathbb{E}_{\varphi}(\varphi) \) and \( \beta^\text{op}_\varphi := \inf_{\varphi \in P} \mathbb{E}_{\varphi}(1 - \varphi) \), respectively.

Definition 3.17 For the above pessimistic errors, the following function is called the pessimistic trade-off function: \( T^\text{pe}(Q(x), Q(x'))(\alpha) := \inf\{\beta^\text{pe}_\varphi : \alpha^\text{pe}_\varphi \leq \alpha\} \). For the above optimistic errors, the following function is called the optimistic trade-off function: \( T^\text{op}(Q(x), Q(x'))(\alpha) := \sup\{\beta^\text{op}_\varphi : \alpha^\text{op}_\varphi \leq \alpha\} \).

The following theorem is another generalization of the well-known result (Theorem 2.4 in [36]) for standard differential privacy.

Theorem 3.18 For any evidential privacy mechanism \( Q \), the following two statements are equivalent:

1. \( Q \) is \( \epsilon \)-WLDP;

2. For any \( \alpha \in [0, 1] \), \( T^\text{pe}(Q(x), Q(x'))(\alpha) \geq f^\text{pe}_\alpha(\alpha) \) and \( T^\text{op}(Q(x), Q(x'))(\alpha) \leq f^\text{op}_\alpha(\alpha) \) where \( f^\text{pe}_\alpha(\alpha) = \max\{1 - \alpha e^\epsilon, 0, e^{-\epsilon}(1 - \alpha)\} \) and \( f^\text{op}_\alpha(\alpha) = \min\{1 - \alpha e^{-\epsilon}, e^\epsilon(1 - \alpha)\} \).

For the composition, the adversary needs to distinguish between \( Q(x) \times Q(x) \) and \( Q(x') \times Q(x') \). Similarly, we can define pessimistic and optimistic type I and II errors: \( \alpha^{2,\text{pe}}_\varphi, \beta^{2,\text{pe}}_\varphi, \alpha^{2,\text{op}}_\varphi \) and \( \beta^{2,\text{op}}_\varphi \). Moreover, for the hypothesis testing problem for the composition, we define the pessimistic and optimistic trade-off functions similarly: \( T^{\text{pe}}_2(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) := \inf\{\beta^{2,\text{pe}}_\varphi : \alpha^{2,\text{pe}}_\varphi \leq \alpha\} \), and \( T^{\text{op}}_2(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) := \sup\{\beta^{2,\text{op}}_\varphi : \alpha^{2,\text{op}}_\varphi \leq \alpha\} \).

Corollary 3.19 For any \( \alpha \in [0, 1] \), \( T^{\text{pe}}_2(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) \geq f^{2,\text{pe}}_\alpha(\alpha) \) and \( T^{\text{op}}_2(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) \leq f^{2,\text{op}}_\alpha(\alpha) \) where \( f^{2,\text{pe}}_\alpha(\alpha) = \max\{1 - \alpha e^{2\epsilon}, -\alpha + \frac{2}{e^{2\epsilon} + 1}, e^{-2\epsilon}(1 - \alpha)\} \) and \( f^{2,\text{op}}_\alpha(\alpha) = \min\{1 - \alpha e^{-2\epsilon}, e^{2\epsilon}(1 - \alpha), -\alpha + \frac{3 - e^{-2\epsilon}}{e^{-2\epsilon} + 1}\} \).

Both Theorem 3.18 and Corollary 3.19 can be visualized as follows:

For simplicity, we consider the above evidential privacy matrix

\[
Q_{2 \times 3} = \begin{pmatrix}
p & q & 1 - p - q \\
q & p & 1 - p - q
\end{pmatrix}
\]

In Definition 3.1, \( 1 - p - q \) quantifies the conditional probability of the third response “I don’t know”. Similarly, in Definition 3.11, \( p \) and \( q \) are the probabilities of telling truthfully and of lying respectively. However, \( 1 - p - q \) measures the probability of unknown response strategy or possible noncompliance. Unlike SLDP, there are only two responses “Yes” and “No” for response mechanism according to WLDP and “I don’t know” is not an option. In order to obtain a Warner-style randomized response
2 × 2 matrix, we redistribute the mass $1 - p - q$ on the unknown part to those masses on “Yes” and “No” and get the following matrix:

$$Q_\lambda = \begin{pmatrix}
    p + \lambda(1 - p - q) & q + (1 - \lambda)(1 - p - q) \\
    q + (1 - \lambda)(1 - p - q) & p + \lambda(1 - p - q)
\end{pmatrix}$$

When $\lambda = 1$, the associated privacy loss is the largest and is the same as according to Definition 3.13. The respondent is most conservative and make the worst-case analysis. On the other hand, when $\lambda = 0$, the associated privacy loss is the smallest. In this case, the respondent is the most optimistic and assumes the best possibility. Similarly, we can obtain the maximum likelihood estimation $\hat{\pi} = \frac{n_1}{n_1 + (1 - \lambda)(1 - p - q)} - \frac{1}{4(2p - 1)^2}$, and show that $\hat{\pi}$ is an unbiased estimate of $\pi$. From Theorem 3.10, we know that, when $\lambda = 0$, the variance $\text{Var}(\hat{\pi}|N_1 + N_2 \neq 0) = \frac{-n}{(n+1)(p+q)-1}$ (Thm. 3.10). In contrast, according to Walley’s semantics, the privacy loss for $Q_{2 \times 3}$ is defined as $\epsilon^S(p, q) = \ln(\frac{p}{q})$ and its accuracy is $\nu^S(p, q) = \text{Var}(\hat{\pi}|N_1 + N_2 \neq 0) = \frac{-(\pi - \frac{1}{2})^2 + \frac{(p-q)^2}{4(2p-1)^2}}{n}$ (Thm. 3.10). In other words, both $\epsilon^W(p, q)$ and $\nu^W(p, q)$ are obtained according to the worst-case analysis from the perspectives of the respondent and adversary respectively. Similarly, we may obtain $\epsilon^O(p, q)$ and $\nu^O(p, q)$, the optimal privacy loss and estimation error among all possible privacy mechanisms $Q_\lambda$. The following Figure 5 illustrates the relationships among the three trade-offs between privacy and accuracy: $(\epsilon^S(p, q), \nu^S(p, q)), (\epsilon^W(p, q), \nu^W(p, q))$ and $(\epsilon^O(p, q), \nu^O(p, q))$. The rectangle shown in the figure consists of exactly the trade-offs between privacy and accuracy for all possible $Q_\lambda$ with $(\epsilon^W(p, q), \nu^W(p, q))$ as the worst and $(\epsilon^O(p, q), \nu^O(p, q))$ as the best.

**Corollary 3.20** $\epsilon^W(p, q)$ is decreasing with respect to $q$ and $\nu^W(p, q)$ is decreasing with respect to $p$. 

![Figure 4: The trade-off between type I and II errors for WLDP](image)
According to the corollary, we may compare two privacy mechanisms \( Q_{2 \times 3} (p, q) \) and \( Q_{2 \times 3} (p', q') \). If \( p \geq p' \) and \( q \geq q' \), then \( \epsilon^W (p, q) \leq \epsilon^W (p', q') \) and \( \nu^W (p, q) \leq \nu^W (p', q') \). In this case, \( Q_{2 \times 3} (p, q) \) is preferred to \( Q_{2 \times 3} (p', q') \). So the trade-off in Walley’s semantics is similar to the minimax estimation for LDP [11].

### 4 Conclusion

To the best of our knowledge, we are the first to explore differential privacy from a different uncertainty perspective than probability theory. The fact that differential privacy is closely related to statistical analysis [13] may explain why there are few research about DP in other uncertainty theories which don’t support a practical statistical analysis. But belief functions are deeply rooted in fiducial inference, an important school in statistics [6, 31, 26, 25]. It is desirable to develop a belief-function theory of differential privacy. The LDP implicitly requires some assumptions about the adversary’s view of belief functions in privacy mechanism. There are many semantics for belief functions. In this paper, we choose Shafer’s semantics as randomly encoded messages [32] and Walley’s interpretation as imprecise-probabilities [33]. Our work in LDP is motivated by the nonresponse and noncompliance issue in randomized response technique in [35, 15] and discrete distribution estimation problem in [20, 19, 34, 17] where the size of the input alphabet is no less than that of the output alphabet. However, since the number of messages (or the size of the powerset of the output set) is usually larger than that of the input set in our LDP mechanisms, MLE is usually different from empirical estimation in this case and their techniques don’t apply here. Moreover, there is a rich literature to address nonresponse in survey research [24] but most of them regard the issue as a missing-data problem and few of them consider the privacy problem. There seems no obvious LDP definitions for coarsening at random because the outputs of coarsening mechanisms at different inputs are different and hence the adversary can easily distinguish these two inputs. It may be interesting to explore the LDPs for contamination models. There are 2 other possible definitions of SLDP in terms of belief functions and plausibility functions: \( e^{-\epsilon} \leq \frac{\text{bel}_{S'} (E)}{\text{bel}_{S} (E)} \leq e^{\epsilon} \) and \( e^{-\epsilon} \leq \frac{\text{pl}_{S'} (E)}{\text{pl}_{S} (E)} \leq e^{\epsilon} \). Lemma 3.2 and the remarks afterwards actually show their relationships. In future versions, we will elaborate these two different definitions and their relations with Definition 3.1.

In this paper we show a binary composition theorem for each definition (Corollaries 3.6 and 3.19). We believe that, for our two definitions SLDP and WLDP, the composition of the hypothesis-
testing trade-off functions \[21, 2\] converges to some (most probably random-set variant) form of Gaussian DP \[9\] according to some central limit theorem (Chapter 3 in \[27\]). In this paper, we took the first step in this direction and showed the effect of the composition of hypothesis-testing trade-off functions (Corollaries 1 and 4). Moreover, we would like to investigate LDP for belief functions from the perspective of respondents (as in \[37\]) and conduct a series of rigorous surveys to show that our new generalized Warner’s mechanism including “don’t know” as an option can indeed increase user’s willingness to participate.

Acknowledgements

The corresponding author wants to thank Professors Arthur Dempster, Xiao-Li Meng and Ruobin Gong for their support during his visiting scholarship at Harvard Statistics Department when this research was initiated. The definition of LDP according to Walley was inspired by an insightful discussion with Professor Xiao-Li Meng. The research is partly supported by NSFC (61732006) and the third author is supported by NSFC (No.61772534).

References

[1] Dakshi Agrawal and Charu C Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 247–255, 2001.

[2] Borja Balle, Gilles Barthe, Marco Gaboardi, Justin Hsu, and Tetsuya Sato. Hypothesis testing interpretations and Rényi differential privacy. In *International Conference on Artificial Intelligence and Statistics*, pages 2496–2506. PMLR, 2020.

[3] Brooke Bullek, Stephanie Garboski, Darakhshan J Mir, and Evan M Peck. Towards understanding differential privacy: When do people trust randomized response technique? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 3833–3837, 2017.

[4] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. “i need a better description”: An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 3037–3052, 2021.

[5] Fabio Cuzzolin. *The Geometry of Uncertainty - The Geometry of Imprecise Probabilities*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer, 2021.

[6] Arthur Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Math. Stat.*, 38:325–339, 1967.

[7] Arthur P. Dempster. The dempster-shafer calculus for statisticians. *Int. J. Approx. Reason.*, 48(2):365–377, 2008.

[8] Thierry Denoeux. Likelihood-based belief function: Justification and some extensions to low-quality data. *Int. J. Approx. Reasoning*, 55(7):1535–1547, 2014.
[9] Jinshuo Dong, Aaron. Roth, and Wenjie. Su. Gaussian differential privacy. *Journal of the Royal Statistical Society: Series B (JRSSB), to appear*, 2021.

[10] John Duchi, Michael I. Jordan, and Martin J. Wainwright. Local privacy and statistical minimax rates. In *FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, pages 429–438. IEEE Computer Society, 2013.

[11] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of American Statistical Association*, 113(521):182–215, 2018.

[12] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings*, volume 3876 of *Lecture Notes in Computer Science*, pages 265–284. Springer, 2006.

[13] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.

[14] Edwin L. Grab and I. Richard Savage. Tables of the expected value of 1/x for positive bernoulli and poisson variables. *Journal of the American Statistical Association*, 49(256):169–177, 1954.

[15] Blair Graeme, Kosuke Imai, and Yang-Yang Zhou. Design and analysis of the randomized response technique. *Journal of the American Statistical Association*, 110(511):1304–1319, 2015.

[16] Naoise Holohan, Douglas J. Leith, and Oliver Mason. Optimal differentially private mechanisms for randomised response. *IEEE Trans. Inf. Forensics Secur.*, 12(11):2726–2735, 2017.

[17] Zhengli Huang and Wenliang Du. Optrr: Optimizing randomized response schemes for privacy-preserving data mining. In *2008 IEEE 24th International Conference on Data Engineering*, pages 705–714. IEEE, 2008.

[18] Peter J. Huber and Volker Strassen. Minimax tests and Neyman-Pearson tests for capacities. *The Annals of Statistics*, 1(2):251–263, 1973.

[19] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48, pages 2436–2444. JMLR.org, 2016.

[20] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *J. Mach. Learn. Res.*, 17:1:1–17:51, 2016.

[21] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. *IEEE Transactions on Information Theory*, 63(6):4037–4049, 2017.

[22] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? In *FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 531–540. IEEE Computer Society, 2008.
[23] Frank Klawonn and Philippe Smets. The dynamic of belief in the transferable belief model and specialization-generalization matrices. In *UAI '92, Stanford, CA, USA, July 17-19, 1992*, pages 130–137, 1992.

[24] Roderick J. A. Little and Donald B. Rubin. *Statistical analysis with missing data*. Wiley, 2002.

[25] Ryan Martin. False confidence, non-additive beliefs, and valid statistical inference. *International Journal of Approximate Reasoning*, 113:39–73, 2019.

[26] Ryan Martin and Chuanhai Liu. *Inferential models: reasoning with uncertainty*, volume 145. CRC Press, 2015.

[27] Ilya Molchanov. *Theory of Random Sets*, volume 87 of *Probability Theory and Stochastic Modelling*. Springer, 2017.

[28] Jack Murtagh and Salil P. Vadhan. The complexity of computing the optimal composition of differential privacy. *Theory Comput.*, 14(1):1–35, 2018.

[29] Kopo M Ramokapane, Gaurav Misra, Jose Such, and Sören Preibusch. Truth or dare: Understanding and predicting how users lie and provide untruthful data online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15, 2021.

[30] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, N.J., 1976.

[31] Glenn Shafer. Belief function and parametric models (with discussion). *J. Roy. Statist. Soc. Ser. B*, 23:322–352, 1982.

[32] Glenn Shafer and Amos Tversky. Languages and designs for probability judgment. *Cogn. Sci.*, 9(3):309–339, 1985.

[33] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, 1990.

[34] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In Engin Kirda and Thomas Ristenpart, editors, *26th USENIX Security Symposium, USENIX Security 2017, Vancouver, BC, Canada, August 16-18, 2017*, pages 729–745. USENIX Association, 2017.

[35] Stanley Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.

[36] Larry Wasserman and Shuheng Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

[37] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. Towards effective differential privacy communication for users’ data sharing decision and comprehension. In *2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020*, pages 392–410. IEEE, 2020.
**Supplementary Materials**

**Proof of Lemma 3.2:** These two propositions follow directly from the inequality (1) and the facts that \( \text{bel}_x(E) = \sum_{E' \subseteq E} m(E') \) and \( \text{pl}_x(E) = \sum_{E' \cap E \neq \emptyset} m(E') \).

**Proof of Lemma 3.3:** The composition follows from the fact that \( \frac{m^{Q_1}_{x,2}(E_1,E_2)}{m^{Q_2}_{x,2}(E_1,E_2)} = \frac{m^{Q_1}_{x,1}(E_1)}{m^{Q_2}_{x,1}(E_1)} \cdot \frac{m^{Q_2}_{x,2}(E_2)}{m^{Q_2}_{x,1}(E_2)} \).

**Proof of Lemma 3.4:** The proof for \( \epsilon \)-SLDP algorithm is similar to that of the post-processing property for probabilistic privacy mechanisms (Proposition 2.1 in [13]). We prove the proposition only for a deterministic function \( f : Y \rightarrow Z \). For any \( x, x' \in X \), \( m^{fQ}_{x}(E) = m^{Q}_{x}(f^{-1}(E)) \leq e^\epsilon m^{Q}_{x'}(f^{-1}(E)) \).

**Proof of Theorem 3.5:** Let’s assume that the rejection region for the hypothesis testing problem is \( R \). Let \( Q \) be \( \epsilon \)-SLDP.

1. Assume that \( \text{pl}_x^Q(R) \leq \alpha \). It follows that, for any \( E \subseteq Y \), \( -\epsilon \leq \frac{m^Q_{x'}(E)}{m^Q_{x}} \leq \epsilon \). We have that \( e^\epsilon \text{bel}_x^Q(R^c) \geq \text{bel}_x^Q(R^c) \geq 1 - \alpha \). So \( \text{bel}_x^Q(R^c) \geq e^{-\epsilon}(1 - \alpha) \). So, \( \text{pl}_x^Q(R) \leq 1 - e^{-\epsilon}(1 - \alpha) \). Moreover, \( \text{pl}_x^Q(R) \leq e^\epsilon \text{pl}_x^Q(R) = e^\epsilon \text{pl}_x^Q(R) \). By putting these together, we obtain that \( \text{pl}_x^Q(R) \leq \min\{e^\epsilon \alpha, 1 - e^{-\epsilon}(1 - \alpha)\} \). This implies that the type 2 error \( \beta_{e^\epsilon} \geq \max\{1 - e^\epsilon \alpha, e^{-\epsilon}(1 - \alpha)\} \).

2. If we assume that \( \text{bel}_x^Q(R) \leq \alpha' \), then \( \text{pl}_x^Q(R^c) \leq 1 - \alpha' \). Since \( e^{-\epsilon} \text{pl}_x^Q(R^c) \leq \text{pl}_x^Q(R^c) \), \( \text{pl}_x^Q(R) \leq e^\epsilon (1 - \alpha') \). It follows that \( \text{bel}_x^Q(R) \geq 1 - e^\epsilon (1 - \alpha') \). Moreover, \( \text{bel}_x^Q(R) \geq e^{-\epsilon} \text{bel}_x^Q(R) \geq \alpha'e^{-\epsilon} \). By putting these together, we have that \( \text{bel}_x^Q(R) \geq \max\{\alpha'e^{-\epsilon}, 1 - e^\epsilon (1 - \alpha')\} \).

In other words, if type I error \( \alpha_{e^\epsilon} \in [l, L] \), then type II error \( \beta_{e^\epsilon} \in [u(L), U(l)] \) where \( u(\alpha) = \max\{e^{-\epsilon}(1 - \alpha), 1 - \alpha e^\epsilon\} \) and \( U(\alpha) = \min\{e^\epsilon (1 - \alpha), 1 - \alpha e^{-\epsilon}\} \). Since the reasonings above are bidirectional, the direction from (2) to (1) can be shown similarly.

**Proof of Corollary 3.6:** Let \( Q_{2 \times 3} \) denote a known row-stochastic matrix as follows:

\[
Q_{2 \times 3} = \begin{pmatrix}
p & q & 1 - p - q \\
q & p & 1 - p - q \\
\end{pmatrix}
\]

where \( p, q \in [0, 1] \). \( Q_{2 \times 3} \) may be regarded as a generalized Warner’s randomized response mechanism where a respondent answers truthfully with probability \( p \), tells a lie with \( q \) and don’t respond or respond “I don’t know” with probability \( 1 - p - q \). Let \( Q_x \) denotes the first row of \( Q \) and \( Q_x^c \) the second row. Now we consider the hypothesis testing problem by distinguishing between \( Q_x \times Q_x \) and \( Q_{x'} \times Q_{x'} \). Let \( m^{Q_2}_{x,2} \) and \( m^{Q_2}_{x',2} \) denote their corresponding mass functions. Their output set is \( \{0, 1, 2\} \).

\[
\begin{array}{c|c|c|c|c|c}
m^{Q_2}_{x,2} & 0 & 1 & 2 & \{0, 1\} & \{1, 2\} & \{0, 1, 2\} \\
p^2 & 2pq & q^2 & 2q(1 - p - q) & 2p(1 - p - q) & (1 - p - q)^2 \\
\end{array}
\]

And

\[
\begin{array}{c|c|c|c|c|c|c}
m^{Q_2}_{x',2} & 0 & 1 & 2 & \{0, 1\} & \{1, 2\} & \{0, 1, 2\} \\
q^2 & 2pq & p^2 & 2p(1 - p - q) & 2q(1 - p - q) & (1 - p - q)^2 \\
\end{array}
\]

Since \( Q \) is \( \epsilon \)-SLDP, we have
\[ e^{-\epsilon} \leq \frac{1-q}{q}, \frac{1-p}{p} \leq e^{\epsilon}. \]

Let \( R \) be a rejection region and assume that \( \alpha' \leq p_0^{Q \times Q}(R) \leq \alpha \). If \((1-p)^2 \leq \alpha\), then \( p^2 \leq e^{2\epsilon} \alpha \). Since \( e^\epsilon \geq 1 \), \( p \geq \frac{2-\epsilon}{e+1} \). It follows that \((2-p)p \leq e^{2\epsilon}(1-p^2)\) and \(1-(1-p)^2 \leq e^{2\epsilon}(1-p^2)\). So we have that \( p^2 \leq 1 - e^{-2\epsilon}(1 - (1-p)^2) \). Since \( 1 - e^{-2\epsilon}(1 - (1-p)^2) \leq 1 - (1-x)e^{-2\epsilon} \), so we have shown that if the type 1 error is \( \leq \alpha \), then the type 2 error is at least \( u^2(\alpha) := \max\{e^{-2\epsilon}(1-\alpha), -\alpha + \frac{2}{e^2+1}, 1 - \alpha e^{2\epsilon}\} \). For the other part, we can show similarly: if the type 1 error is at most \( \alpha \), then the type 2 error is at least \( U^2(\alpha) := \min\{e^{2\epsilon}(1-\alpha), 1 - \alpha e^{-2\epsilon}, -\alpha + \frac{3}{e^2-2}\} \).

**Lemma 4.1** \( \mathbb{E}\left[ \frac{N_1}{N_1 + N_2} \right] = \frac{n_1}{q_1 + q_2} (1 - q_3^n), \) and \( \mathbb{E}\left[ \frac{N_2}{N_1 + N_2} \right] = \frac{n_2}{q_1 + q_2} (1 - q_3^n) \).

**Proof of Lemma 0.1:** We only prove the first part and the proof of the second is similar.

\[
\mathbb{E}\left[ \frac{N_1}{N_1 + N_2} \right] = \sum_{n_1 + n_2 + n_3 = n, n_1 + n_2 \neq 0} \frac{n_1}{n_1 + n_2} \left( \begin{array}{c} n \\n_1 + n_2 \\ n_3 \n_1 \end{array} \right) \left( \begin{array}{c} n_1 + n_2 \\n_1 \\ n_n \end{array} \right) q_1^{n_1} q_2^{n_2} q_3^{n_3}
\]

\[= \sum_{n_1 + n_2 + n_3 = n, n_1 \neq 0} \frac{n_1}{n_1 + n_2} \left( \begin{array}{c} n \\n_3 \n_1 \end{array} \right) \left( \begin{array}{c} n_1 + n_2 \\n_1 \end{array} \right) q_1^{n_1} q_2^{n_2} q_3^{n_3}
\]

\[= q_1 \sum_{n_1 + n_2 + n_3 = n, n_1 \neq 0} \left( \begin{array}{c} n \\n_3 \n_1 \end{array} \right) \left( \begin{array}{c} n_1 + n_2 - 1 \\n_1 - 1 \end{array} \right) q_1^{n_1-1} q_2^{n_2} q_3^{n_3}
\]

\[= q_1 \sum_{1 \leq n_3 \leq n-1} \left( \begin{array}{c} n \\n_3 \n_3 \end{array} \right) q_3^{n_3} \left[ \sum_{n_1 + n_2 = n-n_3, n_1 \neq 0} \left( \begin{array}{c} n_1 + n_2 - 1 \\n_1 - 1 \end{array} \right) q_1^{n_1-1} q_2^{n_2} \right]
\]

\[= q_1 \sum_{1 \leq n_3 \leq n-1} \left( \begin{array}{c} n \\n_3 \n_3 \end{array} \right) q_3^{n_3} (q_1 + q_2)^{n_1+n_2-1}
\]

\[= \frac{q_1}{q_1 + q_2} [(q_1 + q_2 + q_3)^n - q_3^n] \]

\[= \frac{q_1}{q_1 + q_2} (1 - q_3^n) \]

**Proof of Theorem 3.9:** The equality \( \mathbb{E}\left[ \frac{N_2 q - N_3 p}{(N_1 + N_2)(q-p)} \right] = \pi (1 - q_3^n) \) follows directly from the above lemma and the formula for MLE.

**The proof of Theorem 3.10** is much more involved. It needs the following five necessary lemmas which are not stated in the main text. Recall that \( n \) is the sample size. Let \( A = \sum_{0 \leq n_3 < n} \frac{1}{n-n_3} (n_3)(1-q_3)^{n-n_3} q_3^{n_3} \) and \( f(q_1, n_3) = \sum_{n_1+n_2=n-n_3, n_1 \neq 0} n_1 \left( \begin{array}{c} n_1 + n_2 - 1 \\n_1 - 1 \end{array} \right) q_1^{n_1-1} q_2^{n_2} \).

**Lemma 4.2**

\[f(q_1, n_3) = \begin{cases} (q_1 + q_2)^{n-n_3-1} + (n - n_3 - 1)q_1(q_1 + q_2)^{n-n_3-2} & \text{if } n_3 = n - 1, \\
1 & \text{if } n_3 \leq n - 2. \end{cases} \]
\textbf{Proof.} First we compute the integral of $f(x, n_3)$.

$$
\int_0^x f(t, n_3)dt = \sum_{n_1+n_2=n_3, n_1 \neq 0} \int_0^x n_1(n_1+n_2-1) x^{n_1-1} q_2^{n_2}dt
$$

$$
= \sum_{n_1+n_2=n_3, n_1 \neq 0} \left( n_1+n_2-1 \right) x^{n_1} q_2^{n_2}
$$

So $f(x, n_3) = \left( \int_0^x f(t)dt \right)' = (x+q_2)^{n-n_3-1} + (n-n_3-1)x(x+q_2)^{n-n_3-2}$. \text{QED}

**Lemma 4.3** \quad $\sum_{0 \leq n_3 \leq n-1} \frac{1}{n-n_3}(q_1 + q_2)^{n-n_3-1} q_3^{n_3} = \frac{A}{(q_1 + q_2)^2}$.

**Lemma 4.4** \quad $\sum_{0 \leq n_3 \leq n-2} q_1 + q_2)^{n-n_3-2} (q_3^{n_3} = \frac{1-q_3^n-q_3^{-3}}{(q_1 + q_2)^2}$.

**Proof.** It follows from the fact that

$$
[(q_1 + q_2) + q_3] = \sum_{0 \leq n_3 \leq n-2} (q_1 + q_2)^{n-n_3-2} \left( q_3^{n_3} + q_3^n + q_1(q_1 + q_2)q_3^{n-3-1} \right)
$$

\text{QED}

**Lemma 4.5** \quad $\sum_{0 \leq n_3 \leq n-2} (q_1 + q_2)^{n-n_3-2} (q_3^{n_3} = \frac{A-n(q_1 + q_2)q_3^{n-1}}{(q_1 + q_2)^2}$.

**Proof.** It follows directly from the definition of $A$. \text{QED}

**Lemma 4.6** \quad $- \frac{E[(\frac{N_1}{N_1 + N_2})^2]}{E[(\frac{N_2}{N_1 + N_2})^2]} = \frac{-q_1q_2}{(q_1 + q_2)^2} A + \frac{q_1^2}{(q_1 + q_2)^2} (1 - q_3^n)$.

**Proof.** We only show the first part and the proof of the second is similar.

$$
E[(\frac{N_1}{N_1 + N_2})^2] = \sum_{n_1+n_2+n_3=n, n_1 \neq 0} (\frac{n_1}{n_1 + n_2})^2 (\frac{n_1+n_2}{n_1}) (\frac{n_3}{n_3}) q_1 q_2^n q_3^n
$$

$$
= q_1 \sum_{n_1+n_2+n_3=n, n_1 \neq 0} \frac{1}{n-n_3} (\frac{n_1+n_2-1}{n_1-1}) (\frac{n_3}{n_3}) q_1^{n_1-1} q_2^{n_2} q_3^n
$$

$$
= q_1 \sum_{0 \leq n_3 \leq n-1} \frac{1}{n-n_3} \sum_{n_1+n_2+n_3=n, n_1 \neq 0} (\frac{n_1+n_2-1}{n_1-1}) (\frac{n_3}{n_3}) q_1^{n_1-1} q_2^n q_3^n
$$

$$
= q_1 \sum_{0 \leq n_3 \leq n-1} \frac{1}{n-n_3} (q_1 + q_2)^{n-n_3-1} (\frac{n_3}{n_3}) q_3^n + q_1 \sum_{0 \leq n_3 \leq n-2} (q_1 + q_2)^{n-n_3-2} (\frac{n_3}{n_3}) q_3^n
$$

$$
= q_1 \frac{A}{(q_1 + q_2)^2} + q_1 \frac{1}{(q_1 + q_2)^2} (1 - q_3^n)
$$
Proof of Theorem 3:

\[ \text{Var}[\hat{\pi}|N_1 + N_2 \neq 0] = \text{Var}\left[\frac{N_2 q - N_1 p}{(N_1 + N_2)(q - p)}|N_1 + N_2 \neq 0\right] \]

\[ = \text{Var}\left[\frac{N_2 q + N_2 p - (N_1 + N_2)p}{(N_1 + N_2)(q - p)}|N_1 + N_2 \neq 0\right] \]

\[ = \text{Var}\left[\frac{N_2}{N_1 + N_2} \frac{p + q}{q - p} - \frac{p}{q - p} |N_1 + N_2 \neq 0\right] \]

\[ = \frac{(p + q)^2}{(q - p)^2} \text{Var}\left[\frac{N_2}{N_1 + N_2} |N_1 + N_2 \neq 0\right] \]

\[ = \frac{(p + q)^2}{(q - p)^2} \left[ E\left(\frac{N_2}{N_1 + N_2}\right)^2 |N_1 + N_2 \neq 0\right] - E\left[\frac{N_2}{N_1 + N_2} |N_1 + N_2 \neq 0\right]^2 \]

\[ = \frac{(p + q)^2}{(q - p)^2} \left(\frac{q_1 q_2}{(q_1 + q_2)^2} A + \frac{q_2^2}{(q_1 + q_2)^2} - \frac{q_2^2}{(q_1 + q_2)^2}\right) \text{(Lemma 0.6)} \]

\[ = \frac{1}{(q - p)^2} q_1 q_2 A \]

\[ = \frac{1}{(q - p)^2} [\pi p + (1 - \pi)q][\pi q + (1 - \pi)p] A \]

**Proof of Corollary 3.11:** \( f'(q) = \frac{(n+1)(q+3p)+(\pi-1/2)(p+q)}{(p-q)^2(n+1)} > 0. \)

**Proof of Corollary 3.12:** When \( p + q \) is a constant, the optimality of \( Q_{GWRR} \) follows from the same arguments for \( G_{WRR} \) (Example 1) in [33] and [16].

**Proof of Lemma 3.14:** \( P_{bel_x} = \{pr : pr\text{ is a probability distribution and } pr(A) \geq bel_x(A) \text{ for all } A\} \)

and \( P_{bel_{x'}} = \{pr : pr\text{ is a probability distribution and } pr(A) \geq bel_{x'}(A) \text{ for all } A\} \). For a given \( E \subseteq Y \), by using specialization matrices in [23], we can always find a \( pr_x \in P_{bel_x} \) such that \( pr_x(E) = pl_x(E) \) and a \( pr_{x'} \) such that \( pr_{x'}(E) = bel_{x'}(E) \). And Lemma 5 follows directly.

**Proof of Lemma 3.15:** The proof is similar to that for Lemma 2.

**Proof of Lemma 3.16:** According to the well-known weak von-Neumann-Birkhoff Lemma, we only need to prove the proposition only for a deterministic function \( f : Y \rightarrow Z \). The proof for \( \epsilon\)-SLDP algorithm is similar to that of the post-processing property for probabilistic privacy mechanisms. For \( \epsilon\)-WLDL, consider the two probability functions \( p_x \) and \( p_{x'} \) such that

\[ r^Q_W = \frac{p_x(E)}{p_{x'}(E)} \]

for some \( E \). Then

\[ r^{fQ}_W = \max_{x, x' \in X, E \subseteq Z} \frac{p_x f^{-1}(E)}{p_{x'} f^{-1}(E)} \]

\[ \leq \max_{x, x' \in X, E \subseteq Z} \frac{p_x(E)}{p_{x'}(E)} \]

\[ \leq r^Q_W. \]
Proof of Theorem 3.18 Let $R$ be a rejection region. Since $Q$ is $\epsilon$-WLDP, the following inequalities hold:

\[ e^{-\epsilon} \leq \frac{pl_Q^Q(R)}{bel_Q^Q(R)}; \frac{pl_Q^Q(R)}{bel_Q^Q(R)} \leq e^\epsilon \quad (4) \]
\[ e^{-\epsilon} \leq \frac{pl_Q^Q(R^c)}{bel_Q^Q(R^c)}; \frac{pl_Q^Q(R^c)}{bel_Q^Q(R^c)} \leq e^\epsilon \quad (5) \]

1. Assume that $pl_Q^Q(R) \leq \alpha$. It follows that $e^{-\epsilon} bel_Q^Q(R) \leq pl_Q^Q(R) \leq \alpha$. This implies that

\[ bel_Q^Q(R) \leq e^\epsilon \alpha. \]

Moreover, since $bel_Q^Q(R^c) \leq e^\epsilon pl_Q^Q(R^c)$,

\[ 1 - e^\epsilon pl_Q^Q(R^c) \leq 1 - bel_Q^Q(R^c) = pl_Q^Q(R) \leq \alpha. \]

So we have that $pl_Q^Q(R^c) \geq e^{-\epsilon}(1 - \alpha)$. Putting all these together, we obtain that $T^{pe}(Q(x), Q(x'))(\alpha) \geq max\{1 - \alpha e^\epsilon, 0, e^{-\epsilon}(1 - \alpha)\}$.

2. Assume that $pl_Q^Q(R^c) \geq \alpha$. Since $pl_Q^Q(R^c) \leq e^\epsilon bel_Q^Q(R^c) = e^\epsilon(1 - pl_Q^Q(R))$, $pl_Q^Q(1 - \alpha)$.

Moreover, $\alpha \leq pl_Q^Q(R^c) \leq e^\epsilon bel_Q^Q(R^c) = e^\epsilon(1 - pl_Q^Q(R))$. It follows that $pl_Q^Q(R) \leq 1 - e^{-\epsilon} \alpha$.

So $T^{op}(Q(x), Q(x'))(\alpha) \leq min\{1 - \alpha e^{-\epsilon}, e^{-\epsilon}(1 - \alpha)\}$.

The other direction of the equivalence can shown in a similar way. We have finished the proof of Theorem 4.

Proof of Corollary 3.19 The proof is similar to that for Corollary 1. The only things that we need to pay much attention to is the direction of the inequalities.

Proof of Corollary 3.20: From the above analysis, we know that $\epsilon_W(p, q)$ is defined as the privacy loss of the following matrix $Q_1$ in Warner’s model:

\[ Q_1 = \begin{pmatrix} 1 - q & q \\ q & 1 - q \end{pmatrix} \]

So $\epsilon_W(p, q) = ln\left(\frac{1 - q}{q}\right)$ and it is decreasing with respect to $q$. And $\nu_W(p, q)$ is defined as the accuracy of the following matrix $Q_0$ in Warner’s model:

\[ Q_0 = \begin{pmatrix} p & 1 - p \\ 1 - p & p \end{pmatrix} \]

So $\nu_W(p, q) = \frac{-(\pi - 1/2)^2 + \frac{1}{4n}(\frac{1}{p} - \frac{1}{q})^2}{n}$ and it is decreasing with respect to $p$. 

20