Local Differential Privacy for Belief Functions

Qiyu Li[1], Chunlai Zhou[1]*, Biao Qin[1], Zhiqiang Xu[2]

[1] Computer Science Dept., Renmin University of China, Beijing, CHINA
[2] Mohamed 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.

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* (Dwork et al. 2006; Dwork and Roth 2014). 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 (Kasiviswanathan et al. 2008; Kairouz, Oh, and Viswanath 2016; Duchi, Jordan, and Wainwright 2013). 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 (Warner 1965). 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 (Xiong et al. 2020; Cummings, Kaptchuk, and Redmiles 2021; Ramokapane et al. 2021). An effective differential privacy communication can increase data-sharing rates (Xiong et al. 2020).

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 response or say "I don't know" to withhold sensitive information which minimizes the questionable ethical consequences of lying in their eyes (Bullek et al. 2017). 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 mechnisms from a more general Dempster-Shafer perspective by representing uncertainty in privacy mechanisms with *belief functions* (Dempster 1967; Shafer 1976). 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*" (Dempster 2008). 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 ϵ-local differential privacy according to Shafer (ϵ-SLDP) (Definition 1) and the other according to Walley (ϵ-WLDP) (Definition 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 5 and 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 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 (Shafer and Tversky 1985). 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

*Corresponding author.

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 π, the true proportion of the people with the sensitive property. We employ some combinatorial techniques to obtain a formula for the estimation accuracy (Theorem 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 (Walley 1990). 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 (ϵ-SLDP and ϵ-WLDP) and Warner's randomized response mechanism (Figure 5).

Dempster-Shafer Theory

Let Ω 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 Ω 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 $bel(\bigcup_{i=1}^{n} A_i) \geq \sum_{\emptyset \neq I \subseteq \{1,\cdots,n\}} (-1)^{|I|+1} bel(\cap_{i \in I} A_i)$ where $A_i \in \mathcal{A}$ for all $i \in \{1, \cdots, n\}$. 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 (Shafer 1976)). In other words, if $m : \mathcal{A} \rightarrow [0,1]$ is a mass assignment, then it determines a belief function $bel : \mathcal{A} \rightarrow [0,1]$ as follows: $bel(A) = \sum_{B \subseteq A} m(B)$ for all $A \in \mathcal{A}$. Moreover, given a belief function bel, we can obtain its corresponding mass function m as follows: $m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} bel(B)$ 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 $bel(A)$ that an agent commits to A. A subset A with nonzero mass is called a *focal set*. The belief function bel is called *Bayesian* if $m(A) = 0$ for all non-singletons A. The corresponding *plausibility function* $pl_m : 2^{\Omega} \rightarrow [0,1]$ is defined by $pl_m(A) = \sum_{E \cap A \neq \emptyset} m(E)$ for all $A \subseteq \Omega$. Whenever the context is clear, we drop the subscript m. For m, bel and pl, 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 $c_1, \cdots, c_n$, 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 Ω. 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 was A. Notice that $m(\emptyset) = 0$ and that the $m(A)$ sum to one. The quantity $bel(A) = \sum_{B \subseteq A} m(B)$ 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 $bel(A)$ 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 bel, let $\mathcal{P}_{bel}$ denote the set of all probability functions which are consistent with or dominate over bel. In other words, $\mathcal{P}_{bel} = \{pr : pr$ is a probability function on Ω and $pr \geq bel\}$ where $pr \geq bel$ means $pr(E) \geq bel(E)$ for all $E \subseteq \Omega$. Due to lack of information, uncertainty can't be represented by a probability function but by a belief function bel. All consistent probability functions are possible. Whenever enough information is available, we may specify a probability function from $\mathcal{P}_{bel}$ to represent the uncertainty. One may refer to (Cuzzolin 2021) and (Dwork and Roth 2014) for a detailed introduction to belief functions and DP.

Local Differential Privacy

Let X be a private source of information defined on a discrete, finite input alphabet $X = \{x_1, \cdots, x_k\}$ and Y be an output alphabet $Y = \{y_1, \cdots, y_l\}$ 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 $m_x^Q(E)$ (belief $bel_x^Q(E)$ or plausibility $pl_x^Q(E)$) where m_x^Q ($bel_x^Q(E)$ or $pl_x^Q(E)$) is a mass (belief or plausibility) function on Y for all $x \in X$. Since $m_x^Q(\emptyset) = 0$ 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.

LDP according to Shafer

For an evidential privacy mechanism Q, let $r_S^Q = max_{x,x' \in X, E \subseteq Y} \frac{m_x^Q(E)}{m_{x'}^Q(E)}$ and $\epsilon_S^Q = ln(r_S^Q)$.

Definition 1 For any $\epsilon > 0$, the mechanism Q is called ϵ-*locally differential private* according to Shafer (ϵ-SLDP for short) if $-\epsilon \leq \epsilon_S^Q \leq \epsilon$. And ϵ_S^Q is called the *privacy loss* of Q according to Shafer and ϵ 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 ϵ is, the closer the likelihood ratio of $X = x$ to $X = x'$ is to 1. Therefore, when ϵ 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 ϵ-SLDP is just the standard ϵ-LDP for randomized privacy mechanisms. Almost all basic properties for privacy-preserving randomized mechanisms can be generalized to the setting of belief functions. Let $r^Q_{pl,S} = max_{x,x' \in X, E \subseteq Y} \frac{pl^Q_x(E)}{pl^Q_{x'}(E)}$ and $r^Q_{bel,S} = max_{x,x' \in X, E \subseteq Y} \frac{bel^Q_x(E)}{bel^Q_{x'}(E)}$. Denote $\epsilon^Q_{pl,S} := ln(r^Q_{pl,S})$ and $\epsilon^Q_{bel,S} := ln(r^Q_{bel,S})$.

Lemma 2 *If privacy mechanism Q is ϵ-$SLDP$, then $-\epsilon \leq \epsilon^Q_{bel,S} \leq \epsilon$ and $-\epsilon \leq \epsilon^Q_{pl,S} \leq \epsilon$.*

From Lemma 2, we know that $\epsilon^Q_S \geq \epsilon^Q_{pl,S}$. But generally we don't have the converse that $\epsilon^Q_{pl,S} \geq \epsilon^Q_S$. 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 *(Composition) Let Q_1 be an ϵ_1-SLDP evidential privacy mechanism from X to Y_1 and Q_2 be an ϵ_2-SLDP 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 $\epsilon_1 + \epsilon_2$-SLDP.*

The composition of a *data-independent* mapping f with an ϵ locally differential private algorithm Q is also ϵ locally differential private.

Lemma 4 *(Post-processing) Let Q be an ϵ-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 ϵ-SLDP mechanism from X to Z.*

Now we offer a *hypothesis testing* interpretation for the above ϵ-$SLDP$. 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 tradeoff between type I and type II errors (Dong, Roth, and Su 2021). For belief functions, it is natural to consider *minimax tests* (Huber and Strassen 1973). Formally, consider a rejection rule $\phi : Y \to [0, 1]$. Let $\mathcal{P}^Q_x$ and $\mathcal{P}^Q_{x'}$ denote the two sets of probability functions dominating bel^Q_x and $bel^Q_{x'}$ respectively. In other words, $\mathcal{P}^Q_x = \{pr \in \Delta(Y) : pr \geq bel^Q_x\}$ and $\mathcal{P}^Q_{x'} = \{pr \in \Delta(Y) : pr \geq bel^Q_{x'}\}$. The *lower power* of ϕ under x' is defined as $\pi_{x'} := \inf_{pr \in \mathcal{P}^Q_{x'}} \mathbb{E}_{pr}(\phi)$. In the setting of ϵ-$SLDP$, we assume that type I error α_ϕ is represented by $\sup_{pr \in \mathcal{P}^Q_x} \mathbb{E}_{pr}(\phi)$ and type II error by $\beta_\phi = 1 - \inf_{pr \in \mathcal{P}^Q_{x'}} \mathbb{E}_{pr}(\phi)$. A test ϕ is called a *level-α minimax test* if $\phi = argmin\{\beta_\phi : \alpha_\phi \leq \alpha\}$. The following theorem is a generalization of the well-known result (Theorem 2.4 in (Wasserman and Zhou 2010)) for standard differential privacy.

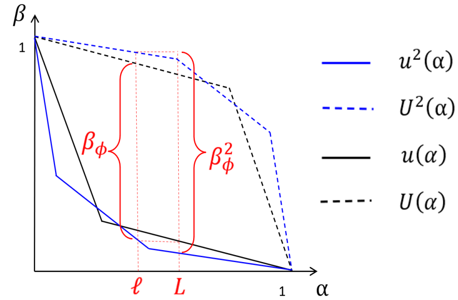


Figure 1: Trade-off between type I and II errors for SLDP

Theorem 5 *For any evidential privacy mechanism Q, the following two statements are equivalent:*

1. *Q is ϵ-SLDP;*
2. *If type I error $\alpha_\phi \in [l, L]$, then type II error $\beta_\phi \in [u(L), U(l)]$ where $u(\alpha) := max\{e^{-\epsilon}(1-\alpha), 1-\alpha e^{\epsilon}\}$ and $U(\alpha) := min\{e^{\epsilon}(1-\alpha), 1-\alpha e^{-\epsilon}\}$.*

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 α^2_ϕ and β^2_ϕ can be defined similarly. For simplicity, we only show the two-fold composition and other multi-fold compositions can be obtained similarly.

Corollary 6 *For the hypothesis testing problem for the composition, if type I error $\alpha^2_\phi \in [l, L]$, then type II error $\beta^2_\phi \in [u^2(L), U^2(l)]$ where $u^2(\alpha) := max\{e^{-2\epsilon}(1-\alpha), -\alpha + \frac{2}{e^\epsilon+1}, 1-\alpha e^{2\epsilon}\}$ and $U^2(\alpha) := min\{e^{2\epsilon}(1-\alpha), 1-\alpha e^{-2\epsilon}, -\alpha + \frac{3-e^{-2\epsilon}}{e^\epsilon+1}\}$.*

Both Theorem 5 and Corollary 6 can be visualized in Figure 1.

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 $\mathbb{S}^k = \{p = (\pi_1, \ldots, \pi_k) : \pi_i \geq 0 (1 \leq i \leq k), \sum_{i=1}^k \pi_i = 1\}$, samples $X_1, \cdots, X_n$ are drawn i.i.d. according to π. 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 π 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 for survey research (Warner 1965).

Example 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 π 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 π as $\hat{\pi} = \frac{p-1}{2p-1} + \frac{n_1}{(p-1)n}$. It can be shown (Warner 1965; Holohan, Leith, and Mason 2017) that this distribution estimation $\hat{\pi}$ is unbiased and its mean square error or variance is the following formula:

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

Within the privacy budget of ϵ, 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 $Q_{2\times3}$ denote a known row-stochastic matrix as follows:

$$Q_{2\times3} = \begin{pmatrix} p & q & 1-p-q \\ q & p & 1-p-q \end{pmatrix}$$

where $p, q \in [0, 1]$. $Q_{2\times3}$ 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 8 In the following we choose to work with such a simple form $Q_{2\times3}$ 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 $Q_{2\times3}$. 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 (Agrawal and Aggarwal 2001). However, the simple form provides us with a neat formula of estimation error (Theorem 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 $Q_{2\times3}$. But we expect that such a simple form to evidential privacy mechanisms is the same as Warner's 2×2 mechanism to the standard LDP. For standard LDP, every approximate DP algorithm can be simulated by a (leaky) variant of Warner's 2×2 mechanism (a well-known result in optimal composition (Murtagh and Vadhan 2018; Kairouz, Oh, and Viswanath 2017)). From a broader and deeper perspective, we believe that every approximate evidential privacy mechanism can be simulated by some variant of our 2×3 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 π 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):

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

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 Yes, the next n_2 say 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 π as follows:

$$\hat{\pi} = \frac{n_2 q - n_1 p}{(n_1 + n_2)(q-p)}. \tag{2}$$

Now we want to compute the expectation of $\hat{\pi}$. From Z_i, we define three new random variables $Z_{i1} = \mathbb{I}_{[Z_i=Yes]}$, $Z_{i2} = \mathbb{I}_{[Z_i=No]}$ 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.

Theorem 9 $\mathbb{E}[\frac{N_2 q - N_1 p}{(N_1+N_2)(q-p)} | N_1 + N_2 \neq 0] = \pi.$

Theorem 10 $Var(\hat{\pi}|N_1 + N_2 \neq 0) = \frac{1}{(q-p)^2}[\pi p + (1-\pi)q][\pi q + (1-\pi)p]A = [-(\pi - \frac{1}{2})^2 + \frac{1}{4}(\frac{p+q}{p-q})^2]A$ *where* $A = \sum_{0 \leq N_3 < n} \frac{1}{n-N_3} \binom{n}{N_3}(1-q_3)^{n-N_3} q_3^{N_3}$.

The formula in Theorem 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 (Grab and Savage 1954) a good approximation of A as $\frac{1}{(n+1)(p+q)-1}$. In particular, with this approximation, when $p+q=1$, $Var[\hat{\pi}|N_1+N_2\neq 0]=\frac{-(\pi-\frac{1}{2})^2+\frac{1}{4}\frac{1}{(2p-1)^2}}{n}$, which is exactly the estimation error of Warner's model (Eq. (1)).

Corollary 11 *Let $f(q)=\frac{-(\pi-\frac{1}{2})^2+\frac{1}{4}(\frac{p+q}{p-q})^2}{(n+1)(p+q)-1}$. Then $f'(q)>0$. In other words, $Var(\hat{\pi})$ is increasing with respect to q.*

This proposition tells us that, within the privacy budget of ϵ, one can increase the estimation accuracy by saying "I don't know" as much as possible instead of lying.

Corollary 12 *Fix $p+q=c$. The optimal ϵ-LDP mechanism is*

$$Q_{GWRR}=\begin{pmatrix} \frac{e^{\epsilon}}{e^{\epsilon}+1}c & \frac{1}{e^{\epsilon}+1}c & 1-c \\ \frac{1}{e^{\epsilon}+1}c & \frac{e^{\epsilon}}{e^{\epsilon}+1}c & 1-c \end{pmatrix}$$

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 $ln(\frac{p}{q})$ as $\epsilon^S(p,q)$ and the estimation error $Var(\hat{\pi}|N_1+N_2\neq 0)$ 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 $e^{\epsilon}=\frac{p}{q}=\frac{p}{1-c-p}$. So we get $p=\frac{1-c}{e^{-\epsilon}+1}$. If we substitute this formula into the error formula in Theorem 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. 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.

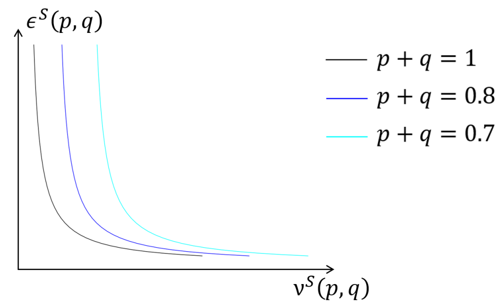


Figure 2: The trade-off in Shafer's semantics

LDP according to Walley

For an evidential privacy mechanism Q, let $r_Q^W = max_{pr_x\in\mathcal{P}_{bel_x^Q}, pr_{x'}\in\mathcal{P}_{bel_{x'}^Q}}\frac{pr_x(E)}{pr_{x'}(E)}$. And the logarithm $\epsilon_Q^W = ln(r_Q^W)$ 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 13 For any $\epsilon>0$, Q is called *ϵ-locally differential private* according to Walley (ϵ-$WLDP$ for short) if, $-\epsilon\leq\epsilon_Q^W\leq\epsilon$. And ϵ_W^Q is called the *privacy loss* of Q according to Walley and ϵ is a *privacy budget*. ◁

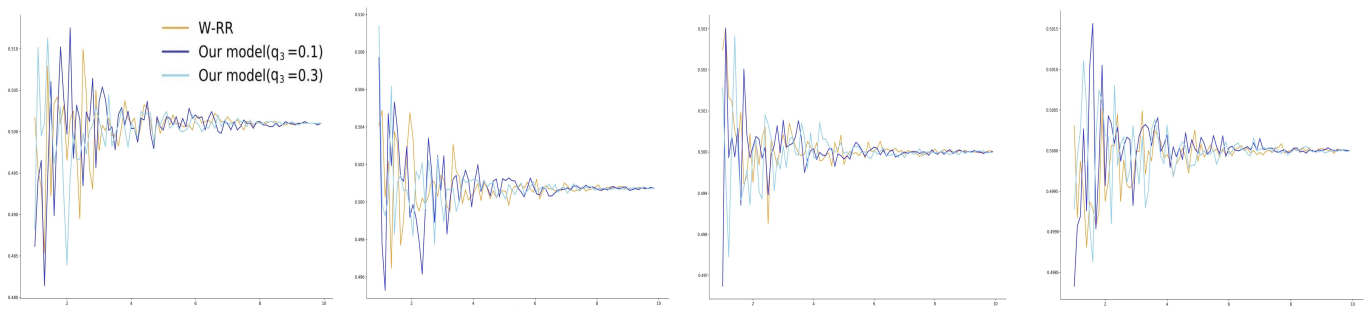
In other words, the privacy loss for ϵ-$WLDP$ is defined by consistent probability functions *in the worst case*. So, ϵ-$WLDP$ fits well with the worst-case analysis behind the philosophy of differential privacy and also with the *conservative* principle of least commitment in the theory of belief functions (Denoeux 2014). Lemma 2 and the following Lemma 14 provide a simple mathematical characterization of SLDP and WLDP, where we can see clearly the main difference between Definitions 1 and 13.

Lemma 14 *(Alternative formulations) If privacy mechanism Q is ϵ-$WLDP$, then, for all $x,x'\in X$ and $E\subseteq Y$: $e^{-\epsilon}\leq\frac{pl_x(E)}{bel_{x'}(E)}\leq e^{\epsilon}$.*

Lemma 15 *(Composition) Let Q_1 be an ϵ_1-WLDP evidential privacy mechanism from X to Y_1 and Q_2 be an ϵ_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 16 *(Post-processing) Let Q be an ϵ-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 ϵ-WLDP mechanism from X to Z.*

For the hypothesis testing problem, recall that Q denotes an evidential privacy mechanism and $\phi: Y\rightarrow[0,1]$ is a rejection rule. In order to translate ϵ-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



(a) N=10 (b) N=100 (c) N=500 (d) N=1000

Figure 3: Impact of sample sizes on the estimation accuracy. The horizontal axis represents the privacy budget ϵ and the vertical axis represents the estimate $\hat{\pi}$.

the other optimistic. For the rejection rule ϕ, the *pessimistic* type I and II are defined as $\alpha_\phi^{pe} = \sup_{pr \in \mathcal{P}_{bel_x^Q}} \mathbb{E}_{pr}(\phi)$ and $\beta_\phi^{pe} = \sup_{pr \in \mathcal{P}_{bel_{x'}^Q}} \mathbb{E}_{pr}(1-\phi)$, respectively. They are actually the same as those errors in ϵ-$SLDP$. Also we define the *optimistic* type I and II errors as $\alpha_\phi^{op} := \inf_{pr \in \mathcal{P}_{bel_x^Q}} \mathbb{E}_{pr}(\phi)$ and $\beta_\phi^{op} := \inf_{pr \in \mathcal{P}_{bel_{x'}^Q}} \mathbb{E}_{pr}(1-\phi)$, respectively.

Definition 17 For the above pessimistic errors, the following function is called the *pessimistic trade-off function*: $T^{pe}(Q(x), Q(x'))(\alpha) := inf\{\beta_\phi^{pe} : \alpha_\phi^{pe} \leq \alpha\}$. For the above optimistic errors, the following function is called the *optimistic trade-off function*: $T^{op}(Q(x), Q(x'))(\alpha) := sup\{\beta_\phi^{op} : \alpha_\phi^{op} \leq \alpha\}$. $\triangleleft$

The following theorem is another generalization of the well-known result (Theorem 2.4 in (Wasserman and Zhou 2010)) for standard differential privacy.

Theorem 18 *For any evidential privacy mechanism Q, the following two statements are equivalent:*

1. *Q is ϵ-WLDP;*
2. *For any $\alpha \in [0,1]$, $T^{pe}(Q(x), Q(x'))(\alpha) \geq f_\epsilon^{pe}(\alpha)$ and $T^{op}(Q(x), Q(x'))(\alpha) \leq f_\epsilon^{op}(\alpha)$ where $f_\epsilon^{pe}(\alpha) = max\{1 - \alpha e^{\epsilon}, 0, e^{-\epsilon}(1-\alpha)\}$ and $f_\epsilon^{op}(\alpha) = min\{1 - \alpha e^{-\epsilon}, e^{\epsilon}(1-\alpha)\}$.*

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: $\alpha_\phi^{2,pe}, \beta_\phi^{2,pe}, \alpha_\phi^{2,op}$ and $\beta_\phi^{2,op}$. Moreover, for the hypothesis testing problem for the composition, we define the pessimistic and optimistic trade-off functions similarly: $T_2^{pe}(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) := inf\{\beta_\phi^{2,pe} : \alpha_\phi^{2,pe} \leq \alpha\}$, and $T_2^{op}(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) := sup\{\beta_\phi^{2,op} : \alpha_\phi^{2,op} \leq \alpha\}$.

Corollary 19 *For any $\alpha \in [0,1]$, $T_2^{pe}(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) \geq f_\epsilon^{2,pe}(\alpha)$ and $T_2^{op}(Q(x) \times Q(x), Q(x') \times Q(x'))(\alpha) \leq f_\epsilon^{2,op}(\alpha)$ where $f_\epsilon^{2,pe}(\alpha) = max\{1 - \alpha e^{2\epsilon}, -\alpha + \frac{2}{e^\epsilon+1}, e^{-2\epsilon}(1-\alpha)\}$ and $f_\epsilon^{2,op}(\alpha) = min\{1 - \alpha e^{-2\epsilon}, e^{2\epsilon}(1-\alpha), -\alpha + \frac{3-e^{-2\epsilon}}{e^\epsilon+1}\}$.*

Both Theorem 18 and Corollary 19 can be visualized in Figure 4.

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 1, $1-p-q$ quantifies the conditional probability of the third response "I don't know". Similarly, in Definition 13, 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 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 $\hat{\pi} = \frac{\frac{n_1}{n} - (1-\lambda)(1-p-q) - q}{p - q + (2\lambda -)(1-p-q)}$, and show that $\hat{\pi}$ is an unbiased estimate of π. From Theorem 10, we know that, when $\lambda = 0$, the variance $Var(\hat{\pi})(= \frac{-(\pi-1/2)^2 + \frac{1}{4(2p-1)^2}}{n})$ is the largest and is defined as *the estimation accuracy* of the privacy matrix $Q_{2\times 3}$ according to Walley.

According to Shafer's semantics, the privacy loss for the mechanism $Q_{2\times 3}$ is defined as $\epsilon^S(p,q) = ln(\frac{p}{q})$ and its accuracy is $\nu^S(p,q) = Var(\hat{\pi}|N_1 + N_2 \neq 0) =$

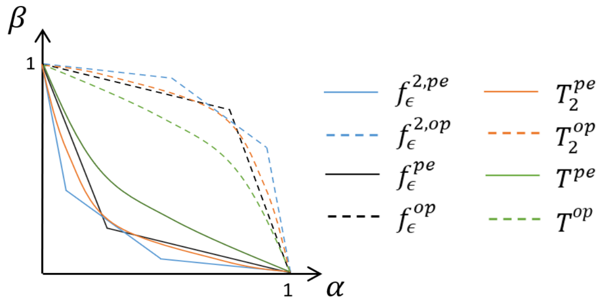


Figure 4: Trade-off between type I and II errors for WLDP

$\frac{-(\pi-\frac{1}{2})^2+\frac{(p-q)^2}{4(p+q)^2}}{(n+1)(p+q)-1}$ (Thm. (10)). In contrast, according to Walley's semantics, the privacy loss for $Q_{2\times3}$ is defined as $ln(\frac{1-q}{q})$, which is denoted as $\epsilon^W(p,q)$ and is equal to the privacy loss of the associated matrix Q_1 in Warner's model. Moreover its accuracy is $\frac{-(\pi-\frac{1}{2})^2+\frac{1}{4(2p-1)^2}}{n}$, which is denoted as $\nu^W(p,q)$ and is exactly the accuracy for the matrix Q_0 in Warner's model. 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_λ. 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_λ 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 20 *$\epsilon^W(p,q)$ is decreasing with respect to q and $\nu^W(p,q)$ is decreasing with respect to p.*

According to the corollary, we may compare two privacy mechanisms $Q_{2\times3}(p,q)$ and $Q_{2\times3}(p',q')$. 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, $Q_{2\times3}(p,q)$ is *preferred* to $Q_{2\times3}(p',q')$. So the trade-off in Walley's semantics is similar to the minimax estimation for LDP (Duchi, Jordan, and Wainwright 2018).

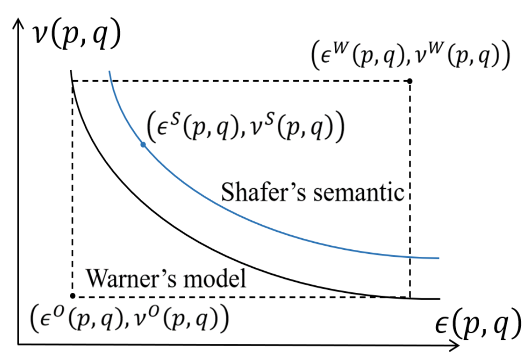


Figure 5: Comparison of trade-offs in the two semantics

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 (Dwork and Roth 2014) 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 (Dempster 1967; Shafer 1982; Martin and Liu 2015; Martin 2019). 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 (Shafer and Tversky 1985) and Walley's interpretation as imprecise-probabilities (Walley 1990). Our work in LDP is motivated by the nonresponse and noncompliance issue in randomized response technique in (Warner 1965; Graeme, Imai, and Zhou 2015) and discrete distribution estimation problem in (Kairouz, Oh, and Viswanath 2016; Kairouz, Bonawitz, and Ramage 2016; Wang et al. 2017; Huang and Du 2008) 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 (Little and Rubin 2002) 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: $e^{-\epsilon}\leq\frac{bel_x^Q(E)}{bel_{x'}^Q(E)}\leq e^\epsilon$ and $e^{-\epsilon}\leq\frac{pl_x^Q(E)}{pl_{x'}^Q(E)}\leq e^\epsilon$. Lemma 2 and the remarks afterwards actually show their relationships. In future versions, we will elaborate these two different definitions and their relations with Definition 1.

In this paper we show a binary composition theorem for each definition (Corollaries 6 and 19). We believe that, for our two definitions SLDP and WLDP, the composition of the hypothesis-testing trade-off functions (Kairouz, Oh, and Viswanath 2017; Balle et al. 2020) converges to some (most probably random-set variant) form of Gaussian DP (Dong, Roth, and Su 2021) according to some central limit theorem (Chapter 3 in (Molchanov 2017)). 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 (Xiong et al. 2020)) 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

Agrawal, D.; and Aggarwal, C. C. 2001. 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*, 247–255.

Balle, B.; Barthe, G.; Gaboardi, M.; Hsu, J.; and Sato, T. 2020. Hypothesis testing interpretations and Rényi differential privacy. In *International Conference on Artificial Intelligence and Statistics*, 2496–2506. PMLR.

Bullek, B.; Garboski, S.; Mir, D. J.; and Peck, E. M. 2017. Towards understanding differential privacy: When do people trust randomized response technique? In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3833–3837.

Cummings, R.; Kaptchuk, G.; and Redmiles, E. M. 2021. " 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*, 3037–3052.

Cuzzolin, F. 2021. *The Geometry of Uncertainty - The Geometry of Imprecise Probabilities*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer. ISBN 978-3-030-63152-9.

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

Dempster, A. P. 2008. The Dempster-Shafer calculus for statisticians. *Int. J. Approx. Reason.*, 48(2): 365–377.

Denoeux, T. 2014. Likelihood-based belief function: Justification and some extensions to low-quality data. *Int. J. Approx. Reasoning*, 55(7): 1535–1547.

Dong, J.; Roth, A.; and Su, W. 2021. Gaussian Differential Privacy. *Journal of the Royal Statistical Society: Series B (JRSSB), to appear*.

Duchi, J.; Jordan, M. I.; and Wainwright, M. J. 2013. Local Privacy and Statistical Minimax Rates. In *FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*, 429–438. IEEE Computer Society.

Duchi, J. C.; Jordan, M. I.; and Wainwright, M. J. 2018. Minimax Optimal Procedures for Locally Private Estimation. *Journal of American Statistical Association*, 113(521): 182–215.

Dwork, C.; McSherry, F.; Nissim, K.; and Smith, A. D. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In Halevi, S.; and Rabin, T., eds., *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*, 265–284. Springer.

Dwork, C.; and Roth, A. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4): 211–407.

Grab, E. L.; and Savage, I. R. 1954. Tables of the Expected Value of 1/X for Positive Bernoulli and Poisson Variables. *Journal of the American Statistical Association*, 49(256): 169–177.

Graeme, B.; Imai, K.; and Zhou, Y.-Y. 2015. Design and Analysis of the Randomized Response Technique. *Journal of the American Statistical Association*, 110(511): 1304–1319.

Holohan, N.; Leith, D. J.; and Mason, O. 2017. Optimal Differentially Private Mechanisms for Randomised Response. *IEEE Trans. Inf. Forensics Secur.*, 12(11): 2726–2735.

Huang, Z.; and Du, W. 2008. OptRR: Optimizing randomized response schemes for privacy-preserving data mining. In *2008 IEEE 24th International Conference on Data Engineering*, 705–714. IEEE.

Huber, P. J.; and Strassen, V. 1973. Minimax tests and Neyman-Pearson tests for capacities. *The Annals of Statistics*, 1(2): 251–263.

Kairouz, P.; Bonawitz, K.; and Ramage, D. 2016. Discrete Distribution Estimation under Local Privacy. In Balcan, M.; and Weinberger, K. Q., eds., *ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48, 2436–2444. JMLR.org.

Kairouz, P.; Oh, S.; and Viswanath, P. 2016. Extremal Mechanisms for Local Differential Privacy. *J. Mach. Learn. Res.*, 17: 17:1–17:51.

Kairouz, P.; Oh, S.; and Viswanath, P. 2017. The Composition Theorem for Differential Privacy. *IEEE Transactions on Information Theory*, 63(6): 4037–4049.

Kasiviswanathan, S. P.; Lee, H. K.; Nissim, K.; Raskhodnikova, S.; and Smith, A. D. 2008. What Can We Learn Privately? In *FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, 531–540. IEEE Computer Society.

Little, R.; and Rubin, D. 2002. *Statistical analysis with missing data*. Wiley. ISBN 9780471183860.

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

Martin, R.; and Liu, C. 2015. *Inferential models: reasoning with uncertainty*, volume 145. CRC Press.

Molchanov, I. 2017. *Theory of Random Sets*, volume 87 of *Probability Theory and Stochastic Modelling*. Springer.

Murtagh, J.; and Vadhan, S. P. 2018. The Complexity of Computing the Optimal Composition of Differential Privacy. *Theory Comput.*, 14(1): 1–35.

Ramokapane, K. M.; Misra, G.; Such, J.; and Preibusch, S. 2021. 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*, 1–15.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton, N.J.: Princeton University Press.

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

Shafer, G.; and Tversky, A. 1985. Languages and Designs for Probability Judgment. *Cogn. Sci.*, 9(3): 309–339.

Walley, P. 1990. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall. ISBN 3-54029586-0.

Wang, T.; Blocki, J.; Li, N.; and Jha, S. 2017. Locally Differentially Private Protocols for Frequency Estimation. In Kirda, E.; and Ristenpart, T., eds., *26th USENIX Security Symposium, USENIX Security 2017, Vancouver, BC, Canada, August 16-18, 2017*, 729–745. USENIX Association.

Warner, S. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association*, 60(309): 63–69.

Wasserman, L.; and Zhou, S. 2010. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489): 375–389.

Xiong, A.; Wang, T.; Li, N.; and Jha, S. 2020. 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*, 392–410. IEEE.