R^2DP: A Universal and Automated Approach to Optimizing the Randomization Mechanisms of Differential Privacy for Utility Metrics with No Known Optimal Distributions

Meisam Mohammady
Concordia University
m_ohamma@encs.concordia.ca

Shangyu Xie
Illinois Institute of Technology
sxie14@hawk.iit.edu

Yuan Hong
Illinois Institute of Technology
yuan.hong@iit.edu

Mengyuan Zhang
Ericsson Research Security
mengyuan.zhang@ericsson.com

Lingyu Wang
Concordia University
wang@ciise.concordia.ca

Makan Pourzandi
Ericsson Research Security
makan.pourzandi@ericsson.com

Mourad Debbabi
Concordia University
debbabi@encs.concordia.ca

ABSTRACT

Differential privacy (DP) has emerged as a de facto standard privacy notion for a wide range of applications. Since the meaning of data utility in different applications may vastly differ, a key challenge is to find the optimal randomization mechanism, i.e., the distribution and its parameters, for a given utility metric. Existing works have identified the optimal distributions in some special cases, while leaving all other utility metrics (e.g., usefulness and graph distance) as open problems. Since existing works mostly rely on manual analysis to examine the search space of all distributions, it would be an expensive process to repeat such efforts for each utility metric. To address such deficiency, we propose a novel approach that can automatically optimize different utility metrics found in diverse applications under a common framework. Our key idea comes from the known fact in probability theory that, by regarding the variance of the injected noise itself as a random variable, a two-fold distribution may approximately cover the search space of all distributions. Therefore, we can automatically find distributions in this search space to optimize different utility metrics in a similar manner, simply by optimizing the parameters of the two-fold distribution. Specifically, we define a universal framework, namely, *r*andomizing the *r*andomization mechanism of *d*ifferential *p*rivacy (R^2DP), and we formally analyze its privacy and utility. Our experiments show that R^2DP can provide better results than the baseline distribution (Laplace) for several utility metrics with no known optimal distributions, whereas our results asymptotically approach to the optimality for utility metrics having known optimal distributions. As a side benefit, the added degree of freedom introduced by the two-fold distribution allows R^2DP to accommodate the preferences of both data owners and recipients.

1 INTRODUCTION

Significant amounts of individual information are being collected and analyzed today through a wide variety of applications across different industries [2]. Differential privacy has been widely recognized as the de facto standard notion [20, 23] in protecting individuals' privacy during such data collection and analysis. On the other hand, since the privacy constraints (e.g., the degree of randomization) imposed by differential privacy may render the released data less useful for analysis, the fundamental trade-off between privacy and utility (i.e., analysis accuracy) has attracted significant attention in various settings [23, 25, 28, 54, 64, 67].

1.1 Motivation

In this context, a key issue is to identify the optimal randomization mechanisms (i.e., distributions and their parameters) [4, 10, 31–33, 35, 38, 41]). While optimizing the parameters of a given distribution can be easily automated, identifying the optimal distribution for different utility metrics is more challenging, and typically requires manual analysis to examine the search space of all distributions. In fact, recent studies [4, 10, 31–33, 35, 38, 41] have only identified the optimal randomization mechanisms for a limited number of cases with specific utility metrics and queries. For instance, Ghosh et al. [35, 38] showed that an optimal randomization mechanism (adding a specific class of *geometric noise*) can be used to preserve differential privacy under the class of negative expected loss utility metrics for a single counting query. Subsequently, Geng et al. [33] showed that, under the ℓ_1 and ℓ_2 norms, the widely used standard Laplace mechanism is asymptotically optimal as $\epsilon \to 0$, whereas the Staircase mechanism (which can be viewed as a geometric mixture of uniform probability distributions) performs exponentially better than the Laplace mechanism in case of weaker privacy guarantees (a comprehensive literature review will be given in Section 6).

However, this has left the optimal distributions of many other utility metrics as open problems, e.g., usefulness (for machine learning applications [7]), entropy-based measures (for signal processing applications [17, 74], and semi-supervised learning [37]), and graph distance metrics (for social network applications [49]). As shown in the works of Ghosh et al. [35, 38] and Geng et al. [33], different utility metrics will likely lead to different optimal distributions. Moreover, since those existing works mostly rely on manual analysis to examine the search space of all distributions, it would be an expensive process to repeat such efforts for each utility metric. Consequently, many existing works simply employ a well-known distribution (e.g., Laplace noise with constant scale parameter or

Gaussian noise with constant variance) without worrying about its optimality. Unfortunately, as our experimental results will show (Section 5), choosing a non-optimal distribution (even with its parameters optimized) may lead to rather poor utility.

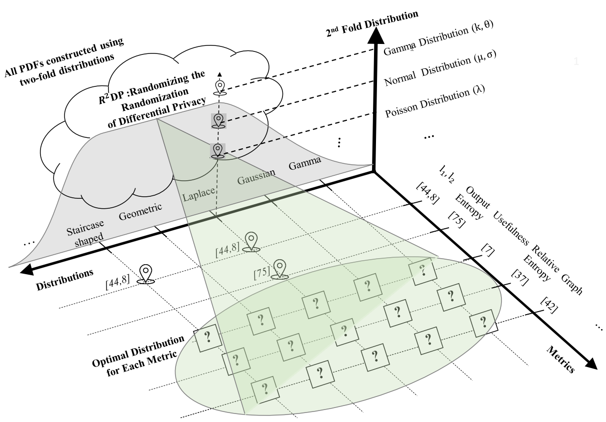


Figure 1: R^2DP can automatically optimize different utility metrics which have no known optimal distributions.

1.2 R^2DP: A Universal Framework

Our key observation is the following. To build a universal framework that can automatically find the optimal distribution in the search space of all distributions, we would need a formulation to link the differential privacy guarantee to the parameters of different distributions (e.g., in Laplace mechanism, ϵ is proportionally related to the inverse of variance). However, it is a known fact that such a formulation varies for each distribution, which explains why existing works have to rely on manual efforts to cover the search space of all distributions, and it also becomes the main obstacle to finding a universal solution that works for all utility metrics employed in different applications.

As depicted in Figure 1, our key idea is that, although it is not possible to directly cover the search space of all distributions in an automated fashion, we can indirectly do so based on the following known fact in probability theory, i.e., a two-fold randomization over the exponential class of distributions may yield many other distributions to approximately cover the search space [16]. Since this class of distributions are all originated from one of the exponential family distributions, their differential privacy guarantee will become a unique function of the parameters of the second fold distribution. Therefore, these parameters can be used to automatically optimize utility w.r.t. different utility metrics through a universal framework, namely, *r*andomizing the *r*andomization mechanism in *d*ifferential *p*rivacy (R^2DP). Furthermore, the two-fold distribution introduces an added degree of freedom, which allows R^2DP to incorporate the requirements of both data owners and data recipients.

1.3 Contributions

Specifically, we make the following contributions:

(1) We define the R^2DP framework with several unique benefits. First, it provides the first universal solution that is applicable to different utility metrics, which makes it an appealing solution for applications whose utility metrics have no known optimal distributions (e.g., [7, 17, 49, 74]). Second, unlike most existing works which rely on manual analysis [35, 38], R^2DP can automatically identify a distribution that yields near-optimal utility, and hence is more practical for emerging applications. Third, R^2DP can incorporate the requirements of both data owners and data recipients, which addresses a practical limitation of most existing approaches, i.e., only the privacy budget ϵ is considered in designing the differentially private mechanisms.

(2) We formally benchmark R^2DP under the well-studied Laplace mechanism. We tackle several key challenges related to the two-fold distribution in R^2DP. We then show that this mechanism yields a class of log-convex distributions for which the differential privacy guarantee can globally be given in terms of the PDFs' parameters. We also show that it can generate near-optimal results w.r.t. a variety of utility metrics whose optimality is known, e.g., Staircase-shape distribution for large ϵ and Laplace itself for small ϵ [33].

(3) We evaluate R^2DP using six different utility metrics, both numerically and experimentally on real data, using both statistical queries (e.g., count and average), and data analytics applications (e.g., machine learning and social network). The experimental results demonstrate that R^2DP can significantly increase the utility for those utility metrics with no known optimal distributions (compared to the baseline Laplace distribution). We also evaluate the optimality of R^2DP using utility metrics whose optimal distributions are known (e.g., Staircase-shape for ℓ_1 and ℓ_2 norms [33]) and our results confirm that R^2DP can generate near-optimal results.

(4) We discuss the potential of adapting R^2DP to improve a variety of other applications related to differential privacy, e.g., query-workload answering.

The rest of the paper is organized as follows. Section 2 provides some related background. Section 3 defines the R^2DP framework. Section 4 formally studies the differential privacy guarantee and the utility of R^2DP. Section 5 presents the experiments. Section 6 reviews the related work, and Section 7 concludes the paper.

2 PRELIMINARIES

We review some background on differential privacy for the theoretical foundations of the R^2DP framework.

2.1 Differential Privacy

We follow the standard definitions of ϵ-differential privacy [25, 64]. Let D be a dataset of interest and d, d' be two adjacent subsets of D meaning that we can obtain d' from d simply by adding or subtracting the data of one individual. A randomization mechanism $\mathcal{M} : \mathrm{D} \times \Omega \rightarrow \mathrm{R}$ which is ϵ-differentially private, necessarily randomizes its output in such a way that for all $S \subset \mathrm{R}$,

$$\mathbb{P}(\mathcal{M}(d) \in S) \leq e^{\epsilon}\mathbb{P}(\mathcal{M}(d') \in S) \tag{1}$$

If the inequality fails, then a leakage (ϵ breach) takes place, which means the difference between the prior distribution and posterior one is tangible. We recall below a basic mechanism that can be used to answer queries in an ϵ-differentially private way. We will only be concerned with queries that return numerical answers, i.e., a query is a mapping $q : \mathrm{D} \to \mathbb{R}$, where $\mathbb{R}$ is a set of real numbers. The following sensitivity concept plays an important role in the design of differentially private mechanisms [23].

Definition 2.1. *The sensitivity of a query* $q : \mathrm{D} \to \mathrm{R}$ *is defined as* $\Delta q = \max_{d,d':Adj(d,d')} |q(d) - q(d')|$ *[25, 64].*

2.2 Laplace Mechanism

The Laplace mechanism [23] modifies an answer to a numerical query by adding zero-mean noise distributed according to a Laplace distribution. Recall that the Laplace distribution with mean zero and scale parameter b, denoted $Lap(b)$, has density $p(x;b) = \frac{1}{2b} exp(-\frac{|x|}{b})$ and variance $2b^2$.

Theorem 2.1. *Let* $q : \mathrm{D} \to \mathbb{R}$ *be a query ,* $\epsilon > 0$*. Then the mechanism* $\mathcal{M}_q : \mathrm{D} \times \Omega \to \mathbb{R}$ *defined by* $\mathcal{M}_q(d) = q(d) + w$*, with* $w \sim Lap(b)$*, where* $b \geq \frac{\Delta q}{\epsilon}$*, is* ϵ*-differentially private [23].*

2.3 Utility Metrics

ℓ_p Metrics. In penalized regression, "ℓ_p penalty" refer to penalizing the ℓ_p norm of a solution's vector of parameter values (i.e., the sum of its absolute values, or its Euclidean length) [69]. In our privacy-utility setting, the ℓ_p utility metric is defined as follows.

Definition 2.2. *(ℓ_p). For a database mechanism* $\mathcal{M}_q(D)$ *the* ℓ_p *utility metric is defined as* $\mathbb{E}(|\mathcal{M}_q(D) - q(D)|^p)^{1/p}$*.*

Usefulness. Following Blum et al. [7], the following utility metric is commonly used for machine learning.

Definition 2.3. *(Usefulness). A mechanism* $\mathcal{M}_q$ *is* (γ, ζ)*-useful if, with probability* $1 - \zeta$*, for any dataset* $d \subseteq \mathrm{D}$*,* $|\mathcal{M}_q(d) - q(d)| \leq \gamma$*.*

Theorem 2.2. *The Laplace Mechanism is* $(\frac{\Delta q}{\epsilon} \ln \frac{1}{\zeta}, \zeta)$*-useful, or equivalently, the Laplace Mechanism is* $(\gamma, e^{\frac{-\gamma}{b(\epsilon)}})$*-useful [13].*

Mallows Metric. The Mallows metric has been applied for evaluating the private estimation of the degree distribution of a social network [42]. It is defined to test if two samples are drawn from the same distribution. Given two random variables X and Y, we have $Mallow(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (|X_i - Y_i|^p)^{1/p}$ (similar to p-norm).

Relative Entropy (Rényi Entropy). The relative entropy, also known as the *Kullback-Leibler (KL)* divergence, measures the distance between two probability distributions [17]. Formally, given two probability distributions $p(x)$ and $q(x)$ over a discrete random variable x, the relative entropy given by $D(p||q)$ is defined as follows: $D(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$. Further generalization came from Rényi [36, 68], who introduced an indexed family of generalized information and divergence measures akin to the Shannon entropy and KL divergence. Rényi introduced the entropy of order α as $I_\alpha(p||q) = \frac{1}{\alpha - 1} \log(\sum_{x \in \mathcal{X}} p(x)^\alpha q(x)^{1-\alpha})$, $\alpha > 0$ and $\alpha \neq 1$.

3 THE R^{2}DP FRAMEWORK

In this section, we define the R^{2}DP framework and its main building block which is the *Utility-maximized PDF* finder.

3.1 Notions and Notations

In probability and statistics, a random variable (RV) that is distributed according to some parameterized PDFs, with (some of) the parameters of that PDFs themselves being random variables, is known as a *mixture* distribution [16] when the underlying RV is discrete (or a *compound* distribution when the RV is continuous). Compound (or mixture) distributions have been applied in many contexts in the literature [66] and arise naturally where a statistical population contains two or more sub-populations.

Definition 3.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* X *be a RV that is distributed according to some parameterized distribution* $f(\theta) \in \mathcal{F}$ *with an unknown parameter* θ *that is again distributed according to some other distribution* g*. The resulting distribution* h *is said to be the distribution that results from compounding* f *with* g*,*

$$h(X) = \int_{\mathbb{R}} f(X|\theta) g(\theta) \, \mathrm{d}\,\theta \tag{2}$$

Then for any Borel subset B *of* $\mathbb{R}$*,*

$$\mathbb{P}(X \in B) = \int_B \int_{\mathbb{R}} f(X|\theta) g(\theta) \, \mathrm{d}\,\theta dX \tag{3}$$

In general, we call any differentially private query answering mechanisms that leverage two-fold probability distribution functions in their randomization, an *R^{2}DP mechanism.*

Definition 3.2. *(R^{2}DP Mechanism). Let* $\mathcal{M}_q(d, u) = q(d) \bigoplus \omega(u)$ *be a mechanism randomizing the answer of a query* q *using a random oracle* $\omega(u)$*, where* u *is the set of parameters (mean, variance, etc.) of the PDF of* ω *and* $\bigoplus$ *stands for the corresponding operator. Denote by* $\mathcal{F}$ *the space of PDFs, we call* $\mathcal{M}_q(d, u)$ *an R^{2}DP mechanism if at least one of the parameters* $u_i \in u$*,* $(i \leq |u|)$ *is/are chosen randomly w.r.t. a specified probability distribution* $f_{u_i} \in \mathcal{F}$*.*

In particular, the *R^{2}DP Laplace mechanism* will modify the answer to a numerical query by adding zero-mean noise distributed according to a compound Laplace distribution with the scale parameter b itself distributed according to some distribution f_b.

Example 3.1. *Suppose that the scale parameter* b *in a Laplace mechanism is randomized as follows:*

$$b = \begin{cases} b_1 & w.p. \quad p, \\ b_2 & w.p. \quad 1 - p. \end{cases}$$

Then, the perturbed result $q(D) + Lap(b)$ *is an example R^{2}DP Laplace mechanism using a Bernoulli distribution.*

Definition 3.3. *Let* $q : \mathrm{D} \to \mathbb{R}$ *be a query and suppose* $f_b \in \mathcal{F}$ *is a probability density function of the scale parameter* b*. Then, the mechanism* $\mathcal{M}_q : \mathrm{D} \times \Omega \to \mathbb{R}$*, defined by* $\mathcal{M}_q(d, b) = q(d) + Lap(b)$ *is an R^{2}DP Laplace mechanism that utilizes PDF* f_b*.*

3.2 The Framework

As shown in Figure 2, R^{2}DP framework include the following steps.

R^{2}DP Computation:

- **Step 1**: The data owner specifies the differential privacy budget ϵ and the data recipient specifies his/her query of interest together with its required utility metric.

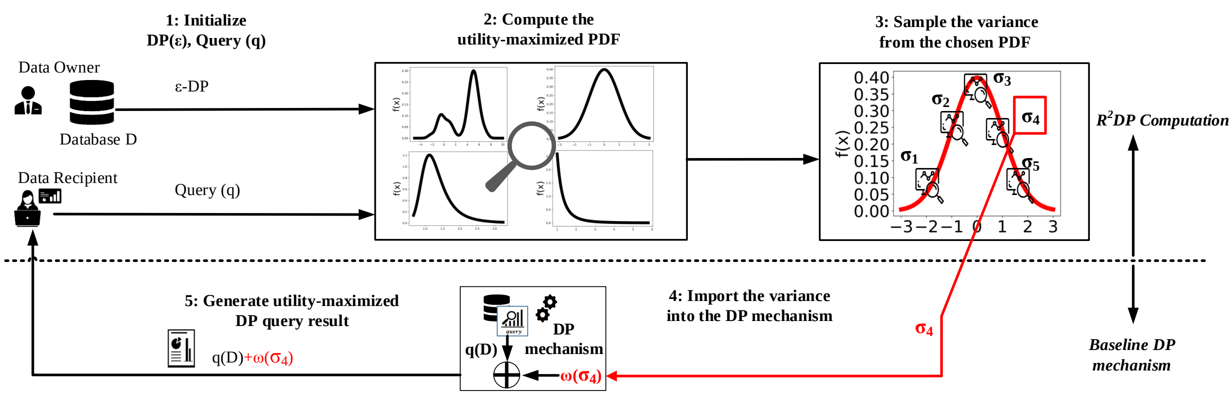


Figure 2: The high level overview of the R^{2}DP framework.

- **Step 2**: Given the input triplets (ϵ, query, metric), the *utility-maximized PDF computing module* computes the provably optimal probability density function and its parameters for the variance of the additive noise. For example, in Figure 2, the PDF computing module returns a lower tail truncated Gaussian distribution for the specified inputs.
- **Step 3**: The *variance sampler* module randomly samples (w.r.t. the PDF found in Step 2) one standard deviation σ_i of the noise to be eventually added.

Baseline DP Randomization:

- **Step 4**: Next, the computed standard deviation σ_i is used to generate a noise $\omega(\sigma_4)$ for the baseline DP mechanism, which is a DP mechanism of exponential order, e.g ., Laplace, Gaussian and exponential mechanisms.
- **Step 5**: The computed noise $\omega(\sigma_i)$ is added to the query result $q(D)$ to provide a utility-maximized DP result to the data recipient.

The most important module of the R^{2}DP framework is the *utility-maximized PDF computing module* (Step 2) which will be described in more details in the following. Furthermore, to make our discussions more concrete, we instantiate the R^{2}DP framework based on the well studied Laplace mechanism, namely, *the R^{2}DP Laplace mechanism*, where other baseline DP mechanisms will be discussed in Appendix F due to space limitation (from now on, we will simply refer to the R^{2}DP Laplace mechanism as R^{2}DP). Particularly, we show that, with a two-fold Laplace distribution, an infinite-size class of log-convex distributions can be identified. This class of distributions pertains a differential privacy guarantee which can globally be given in terms of the PDFs' parameters, and hence is automatically optimizable under the differential privacy constraint.

3.3 Computing Utility-Maximized PDF

In Figure 2, to compute the utility-maximized PDF (Step 2), a key challenge is to establish the search space of automatically optimizable PDFs, from which the utility-maximized PDF is computed. Ideally, the search space of an R^{2}DP mechanism can be defined as the collection of all two-fold distributions, e.g., with Laplace and exponential as the first and second fold distributions, respectively. However, the key challenge here is that a mixture of distributions is itself a distribution which does not necessarily provide a global differential privacy guarantee in terms of the resulting PDFs' parameters (automatically optimizable under the differential privacy constraint). To address this issue, the *Moment Generating Function (MGF)* [30] of the second fold distribution could be utilized, e.g., given the first fold as Laplace distribution. Specifically, MGF of a random variable is an alternative specification of its probability distribution, and hence provides the basis of an alternative route to analytical results compared with directly using probability density functions or cumulative distribution functions [30]. In particular, the MGF of a random variable is a log-convex function of its probability distribution which can provide a global differential privacy guarantee [30] (see Theorem 4.1).

Definition 3.4. *(Moment Generating Function [30]). The moment-generating function of a random variable x is $M_X(t) := \mathbb{E}\left[e^{tX}\right], t \in \mathbb{R}$ wherever this expectation exists. The moment-generating function is the expectation of the random variable e^{tX}.*

Theorem 3.1. *We can write the CDF of the output of an R^{2}DP mechanism in terms of the* Moment Generating Function (MGF) *[30] of the probability distribution $f_{\frac{1}{b}}$, where b is the randomized scale parameter (see Appendix A and C for the details and the proof).*

Thus, for a PDF with non-negative support (since scale parameter is always non-negative), the R^{2}DP mechanism outputs another PDF using the MGF (where CDF is the moment and PDF is its derivative, as shown in Equation 10 in Appendix C) . Moreover, since MGF is a bijective function [29], the R^{2}DP mechanism can in fact generate a search space as large as the space of all PDFs with non-negative support and an existing MGF. However, the next challenge is that not all random variables have moment generating functions (MGFs), e.g., Cauchy distribution [12]. Fortunately, MGFs possess an appealing composability property between independent probability distributions [16], which can be used to provide a search space of all linear combinations of a set of popular distributions with known MGFs (infinite number of RVs).

Theorem 3.2 (MGF of Linear Combination of RVs). *If $x_1, \cdots, x_n$ are n independent RVs with MGFs $M_{x_i}(t) = \mathbb{E}(e^{tx_i})$ for $i = 1, \cdots, n$, then the MGF of the linear combination $Y = \sum_{i=1}^{n} a_i x_i$ is $\prod_{i=1}^{n} M_{x_i}(a_i t)$.*

Consequently, we define the search space of the R^{2}DP mechanism as all possible linear combinations of a set of independent RVs with existing MGF (Section 4.2.2 will provide more details on how to choose the set of independent RVs). Although this search space is only a subset of all two-fold distributions, we will show through both numerical results (in Section E) and experiments with real data (Section 5) that this search space is indeed sufficient to generate near-optimal utility w.r.t. all utility metrics (universality).

4 PRIVACY AND UTILITY

In this section, we analyze the privacy and utility of the R^{2}DP, and then discuss extensions for improving and implementing R^{2}DP.

4.1 Privacy Analysis

We now show the R^{2}DP mechanism provides differential privacy guarantee. By Theorem 3.1, the DP bound of the R^{2}DP is

$$e^{\epsilon} = \max_{\forall S \in \mathrm{R}} \left\{ \frac{-M_{\frac{1}{b}}(-|x-q(d)|)|_{S_{\geq q(d)}} + M_{\frac{1}{b}}(-|x-q(d)|)|_{S_{<q(d)}}}{-M_{\frac{1}{b}}(-|x-q(d')|)|_{S_{\geq q(d')}} + M_{\frac{1}{b}}(-|x-q(d')|)|_{S_{<q(d')}}} \right\}$$

Hence, the value of e^{ϵ} only depends on the distribution of reciprocal of the scale parameter b, i.e., $f_{\frac{1}{b}}$. Moreover, an MGF is positive and log-convex [30] where the latter property is desirable in defining various natural logarithm upper bounds, e.g., DP bound. In the following theorem, our MGF-based formula for the probability $\mathbb{P}(\{q(d) + Lap(b)\} \in S)$ can be easily applied to calculate the differential privacy guarantee (see Appendix C for the proof).

Theorem 4.1. *The R^{2}DP mechanism $\mathcal{M}_q(d, b)$ is*

$$\ln \left[\frac{\mathbb{E}(\frac{1}{b})}{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-\Delta q}} \right] \text{-differentially private.} \tag{4}$$

Moreover, Theorem 3.2 can be directly applied to calculate the differential privacy guarantee of any RV from the search space defined in Section 3.3 (i.e., all linear combinations of a set of independent RVs with known MGFs).

Corollary 4.2 (Differential Privacy of Combined PDFs). *If $x_1, \cdots, x_n$ are n independent random variables with respective MGFs $M_{x_i}(t) = \mathbb{E}(e^{tx_i})$ for $i = 1, \cdots, n$, then the R^{2}DP mechanism $\mathcal{M}_q(d, b)$ where $\frac{1}{b}$ is defined as the linear combination $\frac{1}{b} = \sum_{i=1}^{n} a_i x_i$ is ϵ-differentially private, where*

$$\epsilon = \ln \left[\frac{\sum_{j=1}^{n} a_j \cdot E_{x_j}(\frac{1}{b})}{\sum_{j=1}^{n} a_j \cdot M'_{x_j}(-a_j \cdot \Delta q) \cdot \prod_{\substack{i=1 \\ i \neq j}}^{n} M_{x_i}(-a_i \cdot \Delta q)} \right] \tag{5}$$

Therefore, we have established a search space of probability distributions with a universal formulation for their differential privacy guarantees, which is the key enabler for the universality of R^{2}DP. Next, we characterize the utility of R^{2}DP mechanisms.

4.2 Utility Analysis

We now characterize the utility of the R^{2}DP mechanism. To make concrete discussions, we focus on the usefulness metric (see Section 2), and a similar logic can also be applied to other metrics.

4.2.1 Characterizing the Utility. Denote by $U(\epsilon, \Delta q, \gamma)$ the usefulness of an R^{2}DP mechanism for all $\epsilon > 0$, sensitivity Δq and error bound γ. The optimal usefulness is then given as the answer of the following optimization problem over the search space of PDFs.

$$\max_{f_{\frac{1}{b}} \in F} \{U(\epsilon, \Delta q, \gamma)\} = \max_{f_{\frac{1}{b}} \in F} \left\{ \frac{1}{2} \cdot \left[-M_{\frac{1}{b}}(-|x-q(d)|)|_{q(d)}^{q(d)+\gamma} + M_{\frac{1}{b}}(-|x-q(d)|)|_{q(d)-\gamma}^{q(d)} \right] \right\},$$

$$\text{subject to} \quad \epsilon = \ln \left[\frac{\mathbb{E}(\frac{1}{b})}{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-\Delta q}} \right]$$

where the utility function is the probability of generating ϵ-DP query results within a distance of γ-error (using Theorem 3.1). Note that ϵ and Δq do not directly impact the usefulness but they do so indirectly through the differential privacy constraint. Furthermore, as shown in Theorem 4.1, the differential privacy guarantee ϵ over the established search space is a unique function of the parameters of the second-fold distribution.

Corollary 4.3. *Denote by u, the set of parameters for a probability distribution $f_{\frac{1}{b}}$, and by $M_{f(u)}$ its MGF. Then, the optimal usefulness of an R^{2}DP mechanism utilizing $f_{\frac{1}{b}}$, at each triplet $(\epsilon, \Delta q, \gamma)$ is*

$$U_f(\epsilon, \Delta q, \gamma) = \max_{u \in \mathbb{R}^{|u|}} \left\{ \frac{1}{2} \cdot \left[-M_{f(u)}(-|x-q(d)|)|_{q(d)}^{q(d)+\gamma} + M_{f(u)}(-|x-q(d)|)|_{q(d)-\gamma}^{q(d)} \right] \right\},$$

$$\textit{subject to} \quad \epsilon = \ln \left[\frac{\mathbb{E}(\frac{1}{b})}{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-\Delta q}} \right]$$

Since MGFs are positive and log-convex, with $M(0) = 1$, we have $U_f(\epsilon, \Delta q, \gamma) = 1 - \min_{u \in \mathbb{R}^{|u|}} M_{f(u)}(-\gamma)$. Thus, for usefulness metric, the optimal distribution for ϵ is the one with the minimum MGF evaluated at γ. In particular, for a set of privacy/utility parameters, we can find the optimal PDF using the *Lagrange multiplier* [6]. i.e.,

$$\mathcal{L}(u, \lambda) = M_{f(u)}(-\gamma) + \lambda \cdot \left(\ln \left[\frac{\mathbb{E}(\frac{1}{b})}{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-\Delta q}} \right] - \epsilon\right) \tag{6}$$

Moreover, Theorem 3.2 can be directly applied to design a utility-maximizing R^{2}DP mechanism with a sufficiently large search space (with an infinite number of different random variables).

Corollary 4.4 (Optimal Utility for Combined RVs). *If $x_1, x_2, \cdots, x_n$ are n independent random variables with respective MGFs*

$M_{x_i}(t) = \mathbb{E}(e^{tx_i})$ for $i = 1, 2, \cdots, n$, then for the linear combination $Y = \sum_{i=1}^{n} a_i x_i$, the optimal usefulness (similar relation holds for other metrics) under ϵ-differential privacy constraint is given as

$$U_Y(\epsilon, \Delta q, \gamma) = 1 - \min_{\mathcal{A},\mathcal{U}} \left\{ \prod_{i=1}^{n} M_{x_i}(-a_i \gamma) \right\} \tag{7}$$

subject to

$$\epsilon = \ln \left[\frac{\sum_{j=1}^{n} a_j \cdot E_{x_j}(\frac{1}{b})}{\sum_{j=1}^{n} a_j \cdot M'_{x_j}(a_j \cdot -\Delta q) \cdot \prod_{\substack{i=1 \\ i \neq j}}^{n} M_{x_i}(-a_i \cdot \Delta q)} \right]$$

where $\mathcal{A} = \{a_1, a_2, \cdots, a_n\}$ is the set of the coefficients and $\mathcal{U} = \{u_1, u_2, \cdots, u_n\}$ is the set of parameters of the probability distributions of RVs x_i, $\forall i \leq n$.

Similar to the case of a single RV, we can compute the optimal solution for this optimization problem using the Lagrange multiplier function in Equation 6.

4.2.2 Finding Utility-Maximizing Distributions. Since not all second-fold probability distributions can boost the utility of the baseline Laplace mechanism, leveraging all RVs into our search space would only result in redundant computation by the utility-maximized PDF computing module. Accordingly, in this section, we first derive a necessary condition on the differential privacy guarantee of R^{2}DP to boost the utility of the baseline Laplace mechanism (refer to Appendix C for the proof). Using this necessary condition, we can easily filter out those probability distributions that cannot deliver any utility improvement.

Theorem 4.5. *The utility of R^{2}DP with $\epsilon \geq \ln \left[\mathbb{E}_{\frac{1}{b}} \left(e^{\epsilon(b)} \right) \right]$ is always upper bounded by the utility of the ϵ-differentially private baseline Laplace mechanism. Equivalently, for an R^{2}DP mechanism to boost the utility, the following relation is necessarily true.*

$$e^{\epsilon} = \frac{\mathbb{E}(\frac{1}{b})}{M'_{\frac{1}{b}}(-\Delta q)} < M_{\frac{1}{b}}(\Delta q) \tag{8}$$

We note that $\epsilon = \ln \left[\mathbb{E}_{\frac{1}{b}} \left(e^{\epsilon(b)} \right) \right]$ provides a tight upper bound since it gives the overall e^{ϵ} of an R^{2}DP mechanism as the average of differential privacy leakage. Next, we examine a set of well-known PDFs as second-fold distribution to identify the distribution that offers a significantly improved utility compared with the bound given in Theorem 4.5. Promisingly, our analytic evaluations for *three* of these distributions, i.e., Gamma, uniform and truncated Gaussian distributions demonstrate such a payoff (Appendix B theoretically analyzes several case study PDFs). We note that those chosen distributions are general enough to cover many of other probability distributions (e.g., Exponential, Erlang, and Chi-squared distributions are special cases of Gamma distribution).

4.2.3 Deriving Error Bounds. The error bounds of the R^{2}DP mechanism under some well-known utility metrics are shown in Table 1. The key idea in deriving these results is to calculate the mean of each utility metric over the PDF of RV $1/b$ (which is the linear combination of RVs in multiple PDFs). Specifically, given the error bound $e_L(b)$ for deterministic variance (i.e., Laplace mechanism), the total error bound of an R^{2}DP mechanism will be the mean $\int_0^{\infty} e_L(b) f_b(b) db$. The results shown in Table 1 can be easily applied to optimize those metrics in corresponding applications (e.g., ℓ_1 for private record matching [44], ℓ_2 for location privacy [8], usefulness for machine learning [7], Mallows for social network analysis [42], and relative entropy (with a degree α) for semi-supervised learning [37]).

Table 1: Error bound of R^{2}DP under different metrics

Metric	Dependency to Prior	R^{2}DP Error Bound
ℓ_1	independent	$\int_0^{\infty} M_{\frac{1}{b}}(-x)dx$
ℓ_2	independent	$\sqrt{2 \iint_0^{\infty} M_{\frac{1}{b}}(-u)dudx}$
Usefulness	independent	$1 - M_{\frac{1}{b}}(-\gamma)$
Mallows (p)	dependent	$\left([\sum_{i=1}^{n} \|N_i \sim [-M'_{\frac{1}{b}}(-x)/2]\|^p]/n \right)^{1/p}$
Relative Entropy (α)	dependent	$\frac{\log \sum_{x \in \mathcal{X}}^{n} p(x)^{\alpha} q(x)^{1-\alpha}}{\alpha - 1}$ s.t$(q(x) - p(x)) \sim -M'_{\frac{1}{b}}(-x)/2$

In this context, the ℓ_1, ℓ_2 and usefulness metrics (as defined in Section 2.3) are independent to the prior (i.e., not depending on the distribution of the true results). The metrics will be evaluated based on the deviation between the true and noisy results (which does not change regardless of the prior). On the contrary, some other metrics (e.g., Mallows and relative entropy) depend on the prior distribution of the true results [37, 42]. In such cases, the metrics will be evaluated based on the deviation between the true and noisy results w.r.t. the prior in specific experimental settings (we will discuss those specific priors used in the experiments in Section 5).

In addition to the error bounds given in Table 1, an analyst can derive error bounds for more advanced queries, e.g., those pertaining to learning algorithms [22, 47, 71, 84]. Given the error bound of Laplace mechanism in an application (e.g., Linear SVM [47]), the error bound of the R^{2}DP framework for this application can be derived by taking average of the Laplace's result over the PDF of $\frac{1}{b_r}$. In particular, Table 2 demonstrates the error bounds of R^{2}DP for some learning algorithms (as shown in Section 5, those learning algorithms can benefit from integrating R^{2}DP instead of Laplace).

To derive the error bounds shown in Table 1 and Table 2, the noise parameter(s) and the PDFs used in R^{2}DP can be released to a downstream analyst. This will not cause any privacy leakage because, similar to other differential privacy mechanisms, the privacy protection of R2DP comes from the (first-fold) randomization (whose generated random noises are never disclosed), which will not be affected even if all the noise parameter(s) and the PDFs are disclosed (see Section 4.1 and Appendix C for the formal privacy analysis and proof). We note that, although R^{2}DP replaces the fixed variance of a standard differential privacy mechanism with a random variance, this second-fold randomization is not meant to keep the generated parameters (e.g., the variance) secret, but designed to cover a larger search space (as detailed in Section 3.3).

4.3 R^{2}DP Application

We now present the R^{2}DP algorithm as well as discuss its application to more advanced algorithms.

Table 2: R^2DP compared to Laplace w.r.t. error bounds for learning algorithms

	Linear SVM [47]	Bayesian Inference (statistician) [84]	Robust Linear Regression [22]	Naive Bayes [71]
Laplace	$O(\frac{\log(1/\beta)}{\alpha^2} + \frac{1}{\epsilon\alpha} + \frac{\log(1/\beta)}{\alpha\epsilon})$	$O(mn\log(n))[1 - exp(-\frac{n\epsilon}{2\lvert\mathcal{I}\rvert}]$	$O(n^{-\epsilon logn})$	$O(\frac{1}{n\epsilon})$
R^2DP	$O(\frac{\log(1/\beta)}{\alpha^2} + \mathbb{E}_{\frac{1}{b}}(\frac{b}{\alpha} + \frac{b\log(1/\beta)}{\alpha}))$	$O(mn\log(n))[1 - M_{\frac{1}{b}}(-\frac{n}{2\lvert\mathcal{I}\rvert})]$	$O(\mathbb{E}_{\frac{1}{b}}(n^{-\frac{logn}{b}}))$	$O(\mathbb{E}_{\frac{1}{b}}(\frac{b}{n}))$

4.3.1 The R^2DP Algorithm. Algorithm 1 details an instance of the R^2DP framework using linear combination of three different PDFs. In particular, the algorithm with ϵ-DP finds the best second-fold distribution using the Lagrange multiplier function (see Appendix D) that optimizes the utility metric. Then, the algorithm randomly generates the noise using the two-fold distribution (e.g., first-fold Laplace) and injects it into the query.

Input : Dataset D, Privacy budget ϵ, Query $q(\cdot)$, Metric and its parameters (from data recipient)
Output: Query result $q(D) + Lap(b_r)$, DP guarantee ϵ, Second-fold PDF's parameters

$\Delta q \leftarrow$ Sensitivity $(q(\cdot))$
Find optimal parameters from Lagrange Multiplier $\mathcal{L}(\epsilon, \Delta q, \text{metric}) = a_1^{opt}, a_2^{opt}, a_3^{opt}, k^{opt}, \theta^{opt}, a_u^{opt}, b_u^{opt}, \mu^{opt}, \sigma^{opt}, a_{\mathcal{N}^T}^{opt}, b_{\mathcal{N}^T}^{opt}$
$X_1 \sim \Gamma(k^{opt}, \theta^{opt})$
$X_2 \sim U(a_u^{opt}, b_u^{opt})$
$X_3 \sim \mathcal{N}^T(\mu^{opt}, \sigma^{opt}, a_{\mathcal{N}^T}^{opt}, b_{\mathcal{N}^T}^{opt})$
$\frac{1}{b_r} = a_1^{opt} \cdot X_1 + a_2^{opt} \cdot X_2 + a_3^{opt} \cdot X_3$
return $q(D) + Lap(b_r)$, ϵ, $\mathcal{L}(\epsilon, \Delta q, \text{metric})$

Algorithm 1: The Ensemble R^2DP Algorithm

4.3.2 R^2DP and Advanced Algorithms. We briefly discuss how R^2DP can be applied to improve the utility of advanced algorithms for histogram estimation and query-workload answering (e.g., the popular *matrix mechanism* [59]) (more applications, such as composition and local differential privacy, are discussed in Appendix G). In particular, given a workload (aka. a batch of queries), the matrix mechanism generates a different set of queries, called *strategy queries*, which are answered using a standard Laplace or Gaussian mechanism. The noisy answers to the workload queries can then be derived from the noisy answers to the strategy queries [57]. This two-stage process can result in a correlated noise distribution that preserves differential privacy and also increases utility.

Given a triplet (ϵ, query, metric), R^2DP can be applied to replace the Laplace or Gaussian mechanism for answering the strategy queries of the matrix mechanism. As a result, R^2DP will provide additional improvement in utility (in terms of the TotalError as defined in [57]) over the improvement already provided by the matrix mechanism. More specifically, we compare the total errors of Laplace and R2DP mechanisms in Table 3 for specific workloads of interest (similar to those considered in [57]). These two workloads were analyzed in [57] using two n-sized query strategies, each of which can be envisioned as a recursive partitioning of the domain based on the Haar wavelet [78]. We denote by $f(\epsilon, \Delta q)$ the improvement in the TotalError for applying an R^2DP noise instead of a Laplace noise in the matrix mechanism. For instance, leveraging the results of R^2DP (w.r.t. ℓ_1 or ℓ_2) shown in Section 5.2.2, for a workload of size $n = 6$, at $\epsilon = 2.3$, the improvement for range queries ($\Delta q = 36$) and predicate queries ($\Delta q = 64$) are ~20% and ~10%, respectively.

Table 3: Total error of matrix mechanisms comparison (with R^2DP vs. Laplace) – two workloads and two query strategies

TotalError		Matrix Strategies	
Mechanisms	Workload Queries	Binary Hierarchy of Sums	Matrix of the Haar Wavelet
Laplace	Range Queries	$\Theta(n^2\log^3(n)/\epsilon^2)$	$\Theta(n^2\log^3(n)/\epsilon^2)$
	Predicate Queries	$\Theta(n2^n\log^2(n)/\epsilon^2)$	$\Theta(n2^n\log^2(n)/\epsilon^2)$
R^2DP	Range Queries	$\Theta(f(\epsilon, n^2)n^2\log^3(n)/\epsilon^2)$	$\Theta(f(\epsilon, n^2)n^2\epsilon^2\log^3(n))$
	Predicate Queries	$\Theta(f(\epsilon, 2^n)n2^n\log^2(n)/\epsilon^2)$	$\Theta(f(\epsilon, 2^n)n2^n\log^2(n)/\epsilon^2)$

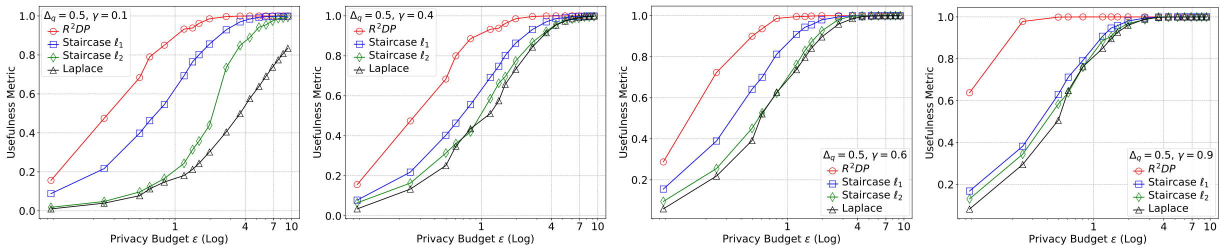
5 EXPERIMENTAL EVALUATIONS

In this section, we experimentally evaluate the performance of the R^2DP framework using six different utility metrics, i.e., ℓ_1, ℓ_2, entropy, usefulness, Mallows and Rényi divergence. Furthermore, we investigate the tightness of the R^2DP mechanism under Rényi differential privacy (RDP in short) [62] which provides a universal formulation of the privacy losses of various DP mechanisms, as shown in Appendix F.2.2 (facilitating the comparison between different mechanisms). Our objective is to verify the following two properties about the performance of the R^2DP framework w.r.t. all seven utility and privacy metrics: (1) R^2DP produces near-optimal results and (2) R^2DP performs strictly better than well-known baseline mechanisms, e.g, Laplace and Staircase mechanisms, in settings where an optimal PDF is not known, e.g., usefulness utility metric or Rényi differential privacy .

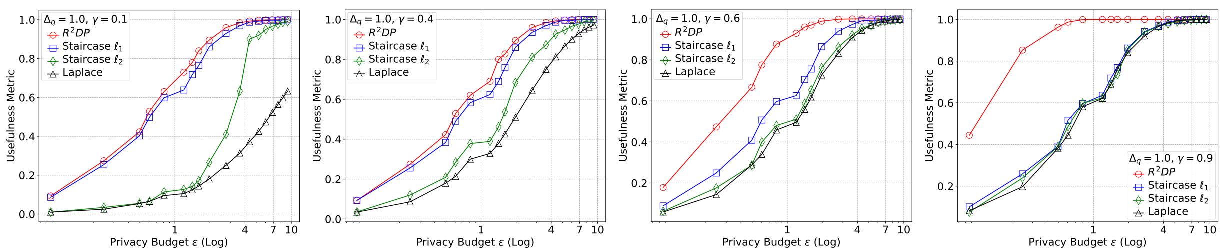
5.1 Experimental Setting

We perform all the experiments and comparisons on the Privacy Integrated Queries (PINQ) platform [61]. Besides basic statistical queries, two applications in the current suite (*machine learning* and *social network analysis*) are employed to evaluate the accuracy of R^2DP and compare it to Laplace and Staircase mechanisms.

5.1.1 Statistical Queries. In the first set of our experiments, we examine the benefits of R^2DP using basic statistical functions, i.e., count and average. The dataset comes from a sensor network experiment carried out in the Mitsubishi Electric Research Laboratories (MERL) and described in [77]. MERL has collected motion sensor data from a network of over 200 sensors for a year and the dataset contains over 30 million raw motion records. To illustrate the query performance with different sensitivities, we create the queries based on a subset of the data including aggregated events that are recorded by closely located sensors over 5-minute intervals. We formed in this way 10 input signals corresponding to 10 spatial zones (each zone is covered by a group of sensors). Since each individual can activate several sensors and travel through different



(a) $\Delta q = 0.5, \gamma = 0.1$ (b) $\Delta q = 0.5, \gamma = 0.4$ (c) $\Delta q = 0.5, \gamma = 0.6$ (d) $\Delta q = 0.5, \gamma = 0.9$



(e) $\Delta q = 1, \gamma = 0.1$ (f) $\Delta q = 1, \gamma = 0.4$ (g) $\Delta q = 1, \gamma = 0.6$ (h) $\Delta q = 1, \gamma = 0.9$

Figure 3: Usefulness metric: R^{2}DP (with five PDFs, i.e., Gamma, Uniform, Truncated Gaussian, Noncentral Chi-squared and Rayleigh distributions) strictly outperforms Laplace and Staircase mechanisms for statistical queries, where the ratio of improvement depends on the values of Δq, γ and ϵ.

zones, we define moving average functions with arbitrary sensitivity values, e.g., $\Delta q \in [0.1, 5]$. For instance, we could be interested in the summation of the moving averages over the past 30 min for zones 1 to 4. We apply R^{2}DP w.r.t. usefulness, ℓ_1, ℓ_2, entropy, and Rényi metrics, respectively.

5.1.2 Social Network. Social network degree distribution is performed on a Facebook dataset [55]. They consist of "circles" and "friends lists" from Facebook by representing different individuals as nodes (47,538 nodes) and friend connections as edges (222,887 edges). Recall that the Mallows metric is frequently used for social network (graph-based) applications [49]. We thus apply R^{2}DP w.r.t. the Mallows metric in this group of experiments.

5.1.3 Machine Learning. Naive Bayes classification is performed on two datasets: Adult dataset (in the UCI ML Repository) [51] and KDDCup99 dataset [70]. First, the Adult dataset includes the demographic information of 48,842 different adults in the US (14 features). It can be utilized to train a Naive Bayes classifier to predict if any adult's annual salary is greater than 50k or not. Second, the KDD competition dataset was utilized to build a network intrusion detector (given 24 training attack types) by classifying "bad" connections and "good" connections. Recall that the usefulness metric is commonly used for machine learning [7]. We thus apply R^{2}DP w.r.t. the usefulness in this group of experiments.

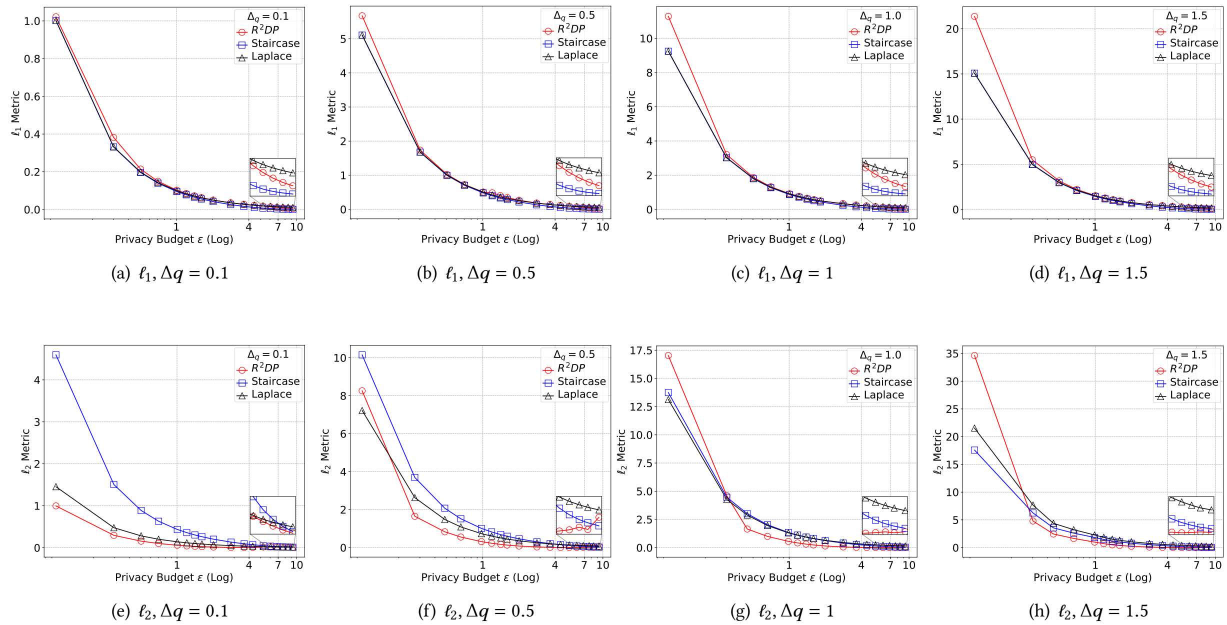
5.2 Basic Statistical Queries

We validate the effectiveness of R^{2}DP using two basic statistical queries: count (sensitivity=1) and moving average with different window sizes, e.g., sensitivity $\in [0.1, 2]$ to comprehensively study the performance of R^{2}DP by benchmarking with Laplace and Staircase mechanisms. We have the following observations.

5.2.1 Usefulness Metric. We compare R^{2}DP with the baseline Laplace and two classes of Staircase mechanisms proposed in [38] w.r.t. ℓ_1 and ℓ_2 metrics, by varying the privacy budget ϵ, four error bounds $\gamma \in \{0.1, 0.4, 0.6, 0.9\}$ and two different sensitivities (Section E additionally shows numerical results to provide a more comprehensive evaluation for the usefulness metric). As shown in Figure 3, R^{2}DP generates strictly better results w.r.t. the usefulness metric, and the ratio of improvement depends on values of ϵ, Δq and γ. In particular, we observe that the improvement is relatively larger for a larger error bound and smaller sensitivity (Figure 3 (a,b,e,f) vs. (c,d,g,h)). One important factor determining the improvement is the ratio between γ and Δq, since it exponentially affects the search space of the R^{2}DP mechanism. Furthermore, we observe that the Laplace and the staircase mechanisms are not optimal (w.r.t. usefulness) for very small and large values of ϵ, respectively, even though they are known to be optimal under other utility metrics (e.g., [34]).

5.2.2 ℓ_1 and ℓ_2 Metrics. We compare R^{2}DP with the baseline Laplace and Staircase mechanisms [38], by varying the privacy budget ϵ and for four different sensitivities $\Delta q \in \{0.1, 0.5, 1, 1.5\}$. Our results validate the findings of Geng et al. [34], i.e., in the low privacy regime ($\epsilon \to \infty$), the Staircase mechanism is optimal while in the high privacy regime ($\epsilon \to 0$), the Laplace mechanism is optimal.

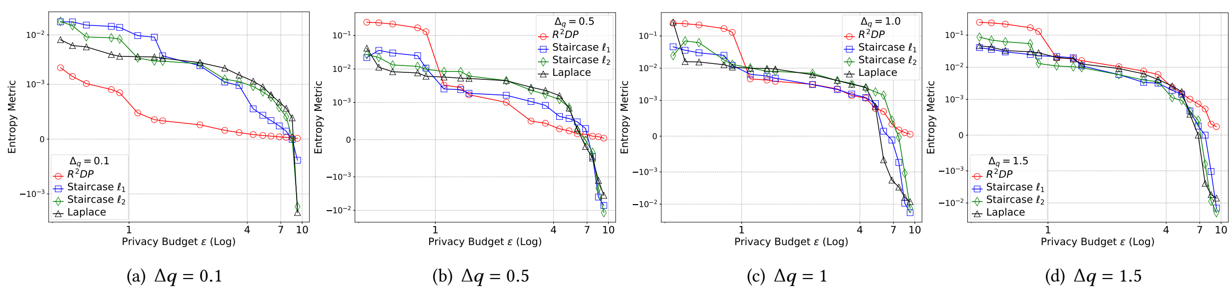
More importantly, our evaluations show that, for medium regime of privacy budgets (which could be more desirable in practice), the class of optimal noise can be totally different. In fact, as shown



(a) $\ell_1, \Delta q = 0.1$ (b) $\ell_1, \Delta q = 0.5$ (c) $\ell_1, \Delta q = 1$ (d) $\ell_1, \Delta q = 1.5$

(e) $\ell_2, \Delta q = 0.1$ (f) $\ell_2, \Delta q = 0.5$ (g) $\ell_2, \Delta q = 1$ (h) $\ell_2, \Delta q = 1.5$

Figure 4: ℓ_1 and ℓ_2 metrics: R^{2}DP compared to Laplace and Staircase mechanisms for statistical queries (with five PDFs, i.e., Gamma, Uniform, Truncated Gaussian, Noncentral Chi-squared and Rayleigh distributions).

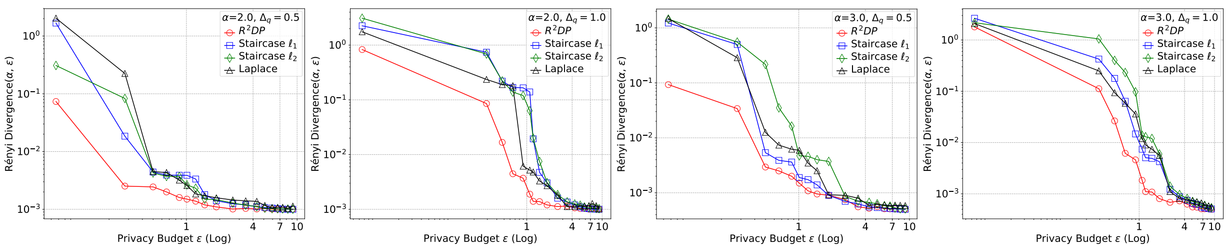


(a) $\Delta q = 0.1$ (b) $\Delta q = 0.5$ (c) $\Delta q = 1$ (d) $\Delta q = 1.5$

Figure 5: KL Divergence (Relative entropy metric): R^{2}DP (with five PDFs, i.e., Gamma, Uniform, Truncated Gaussian, Noncentral Chi-squared and Rayleigh distributions) compared to Laplace and Staircase mechanisms.

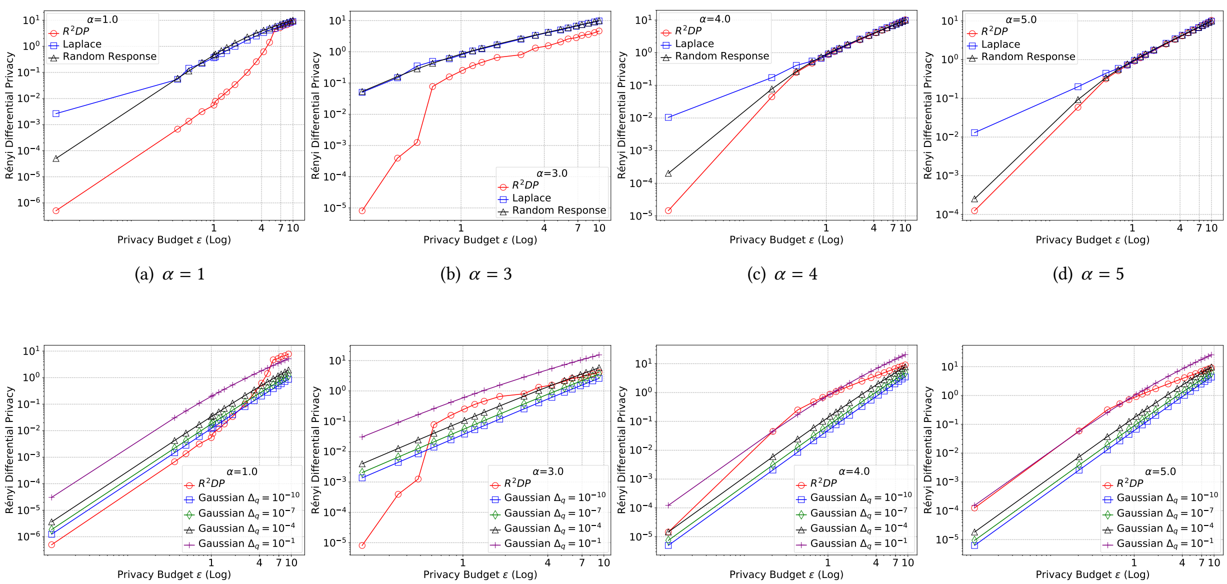
in Geng et al. [34], the lower-bound of ϵ at which the Staircase distribution performs better than the Laplace distribution is somewhere around $\epsilon = 3$ for both ℓ_1 and ℓ_2 metrics. As illustrated in Figure 4, in contrast to ℓ_1 metric (for which the results of laplace and staircase are relatively tight), R^{2}DP can find a class of noises with significantly improved ℓ_2 metric for $\epsilon < 3$ (a logarithmic X axis is used to illustrate the performance in this region). The PDF of this class of noises is mostly two-fold distributions with Laplace distribution as the first fold, and Gamma distribution as the second fold. This finding is in line with the optimal class of noise proposed by Koufogiannis et al. [52], i.e., $f(v) = \dfrac{\epsilon^n \Gamma(\frac{n}{2}+1)}{\pi^{\frac{n}{2}}\Gamma(n+1)} e^{-\epsilon ||v||_2}$. Furthermore, our results suggest different classes of optimal noises (than those found in the literature) for different parameters, sensitivity, ϵ and p (index of ℓ norm). In particular, a larger p tends to provide larger search spaces for R^{2}DP optimization, which results in further improved results for $\epsilon < 3$ (Figure 4 (a,b,e,f) vs. (c,d,g,h)).

5.2.3 Relative Entropy Metric. As Wang et al [75] has already shown that the output entropy of ϵ-DP randomization mechanisms is lower bounded by $1 - \ln(\epsilon/2)$ (for count queries) and the optimal result is achieved with Laplace mechanism, we focus our entropy metric evaluation on relative entropy metrics, i.e., KL and Rényi divergences. To define the prior distribution for this group of experiments, we have created a histogram with 50 bins of our data



(a) $\alpha = 2, \Delta q = 0.5$ (b) $\alpha = 2, \Delta q = 1$ (c) $\alpha = 3, \Delta q = 0.5$ (d) $\alpha = 2, \Delta q = 1$

Figure 6: Rényi Divergence (Relative entropy metric): R^{2}DP (with five PDFs, i.e., Gamma, Uniform, Truncated Gaussian, Non-central Chi-squared and Rayleigh distributions) compared to Laplace and Staircase mechanisms.



(a) $\alpha = 1$ (b) $\alpha = 3$ (c) $\alpha = 4$ (d) $\alpha = 5$

(e) $\alpha = 1$ (f) $\alpha = 3$ (g) $\alpha = 4$ (h) $\alpha = 5$

Figure 7: Rényi Differential Privacy: (a-d) R^{2}DP compared to Laplace and Random Response mechanisms, and (e-h) R^{2}DP compared to Gaussian mechanism.

and calculated the probability mass function (pmf) of the bins. [1] As illustrated in Figure 5, we can draw similar observations for the KL entropy metric. In particular, we observe that R^{2}DP performs better for smaller sensitivity due to the larger search space of PDFs used in optimization. Similarly the Rényi divergence entropy depicted in Figure 6 illustrates a similar trend with given different α (the index of the divergence).

Summary. R^{2}DP mechanism can generate better results than most of the well-known distributions for utility metrics without known optimal distributions (e.g., usefulness), and our results asymptotically approach to the optimal for utility metrics with known optimal distributions (e.g., ℓ_1 and ℓ_2). In particular, even though R^{2}DP is not specifically designed to optimize ℓ_1 and ℓ_2 metrics, we observe a very similar performance between the R^{2}DP results and the optimal Staircase results, e.g., the multiplicative gain compared to the Laplace results. We note that using a larger number of independent RVs drawm from different PDFs as the search space generator may further improve the results.

[1] 2 millions records fall into 50 bins (e.g., equal range for each bin). Then, any counting and moving average query (with different sensitivities) can be performed within each of the 50 bins to generate the distribution. Finally, the distance between the original and noisy distributions can be measured using the relative entropy metrics.

5.3 Tightness of R^2DP under Rényi DP

Rényi differential privacy [62] is a recently proposed as a relaxed notion of DP which effectively quantifies the bad outcomes in (ϵ, δ)-DP mechanisms and consequently evaluates how such mechanisms behave under sequential compositions (see Appendix F.2.2 for details on Rényi DP). We now evaluate how the privacy loss of R^2DP behaves under Rényi DP.

Specifically, this group of experiments are conducted to provide insights about the privacy loss of R^2DP and other well-known mechanisms. In particular, Figure 7 (a-d) depicts the Rényi differential privacy of the R^2DP and two basic mechanisms for counting queries: random response and Laplace mechanisms. These results are based on the privacy guarantees depicted in Table 4. Our results demonstrate that fine tuning R^2DP can generate strictly more private results compared to the other two ϵ-DP mechanisms when the definition of the privacy notion is relaxed. Furthermore, the level of such tightness depends on the Rényi differential privacy index where a smaller value of α pertains to a relatively tighter R^2DP mechanism. On the other hand, all three mechanisms behave more similarly as α increases. Ultimately, at $\alpha \longrightarrow \infty$, where Rényi differential privacy becomes equivalent to the classic notion of ϵ-DP, all three mechanisms' privacy guarantees converge to ϵ.

Table 4: Summary of Rényi DP parameters for four mechanisms based on Theorem F.4

Mechanism	Differential Privacy	Rényi Differential Privacy for α
Laplace	$\frac{1}{b}$	$\alpha > 1 : \frac{1}{\alpha-1} \log \left[\frac{\alpha \cdot exp(\frac{\alpha-1}{b}) + (\alpha-1) \cdot exp(\frac{-\alpha}{b})}{2\alpha-1} \right]$ $\alpha = 1 : \frac{1}{b} + exp(\frac{-1}{b}) - 1$
Random Response	$\lvert \log \frac{p}{1-p} \rvert$	$\alpha > 1 : \frac{1}{\alpha-1} \log \left[p^{\alpha}(1-p)^{1-\alpha} + p^{1-\alpha}(1-p)^{\alpha} \right]$ $\alpha = 1 : (2p-1) \log \frac{p}{1-p}$
R^2DP	$M'_{\frac{1}{b}}(0) / M'_{\frac{1}{b}}(-1)$	$\alpha > 1 : \frac{1}{\alpha-1} \log \left[\frac{\alpha M_{\frac{1}{b}}(\alpha-1) + (\alpha-1) M_{\frac{1}{b}}(-\alpha)}{2\alpha-1} \right]$ $\alpha = 1 : M'_{\frac{1}{b}}(0) + M_{\frac{1}{b}}(-1) - 1$
Gaussian	∞	$\frac{\alpha}{2\sigma^2}$

In the next set of experiments, we compare the R^2DP mechanism and Gaussian mechanism in terms of privacy guarantee to understand how exactly the bad outcomes probability (δ) affects the privacy robustness of a privatized mechanism. Figure 7 (e-h) gives such a comparison. Specifically, since Rényi differential privacy at each α can be seen as higher-order moments as a way of bounding the tails of the privacy loss variable [62], we observe that each value of α reveals a snapshot of such a privacy loss. As a tangible observation, we conclude that the class of optimal ϵ-differential privacy mechanisms benefits from a very smaller privacy loss at smaller moments (which are more decisive in overall protection) and larger privacy loss at bigger moments.

5.4 Social Network Analysis

We conduct experiments to compare the performance of R^2DP, Laplace and two staircase mechanisms based on PINQ queries in social network analysis. Figure 8 compares the degree distribution for a real Facebook dataset using Mallows metric (the prior, i.e.,$n = 47,538$ nodes, and $p = 1$ or 2 for computing the distribution distance using Mallows metric). Again, our results confirm that R^2DP can effectively generate PDFs to maximize this utility metric suitable for social networking analysis. We note that, since the definition of this metric is similar to ℓ_p metric (Mallows is more empirical, depending on the number of nodes in the dataset), the results for this metric display a similar pattern to those for ℓ_p metric depicted in Figure 4.

5.5 Machine Learning

We obtain our baseline results by applying the Naive Bayes classifier on the Adult dataset (45K training records and 5K testing records), the precision and recall results are derived as 0.814 and 0.825, respectively. Then, we evaluate the precision and recall of R^2DP and Laplace-based naive classification [72] by varying the privacy budget for each PINQ query $\epsilon \in [0.1, 10]$ (sensitivity=1) where two different error bounds $\gamma = 0.05, 0.1$ are specified for R^2DP. We have the following observations:

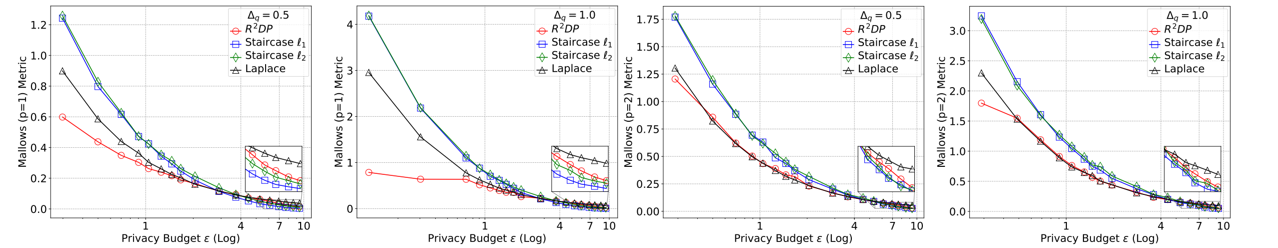
- As shown in Figure 9(a) and 9(b), the R^2DP-based classification is more accurate than the Laplace and staircase mechanisms with the same total privacy budget for all the PINQ queries ϵ. As the privacy budget ϵ increases, following our statistical query experiments, R^2DP offers a far better precision/recall compared to the Laplace-based classification (close to the results without privacy consideration) since it approaches to the optimal PDF.
- Among the precision/recall results derived with two different γ in R^2DP-based classification, for each ϵ, one out of the two specified error bounds (e.g., $\gamma = 5\%$) may reach the highest accuracy (not necessarily the result with the smaller γ).
- As shown in Figure 9(c) and 9(d), we can draw similar observations from the KDDCup99 dataset.

The above experimental results have validated the effectiveness of integrating R^2DP to improve the output utility for classification while ensuring ϵ-differential privacy. In summary, all the experiments conducted in both statistical queries and real-world applications have validated the practicality of the R^2DP framework.

6 RELATED WORK

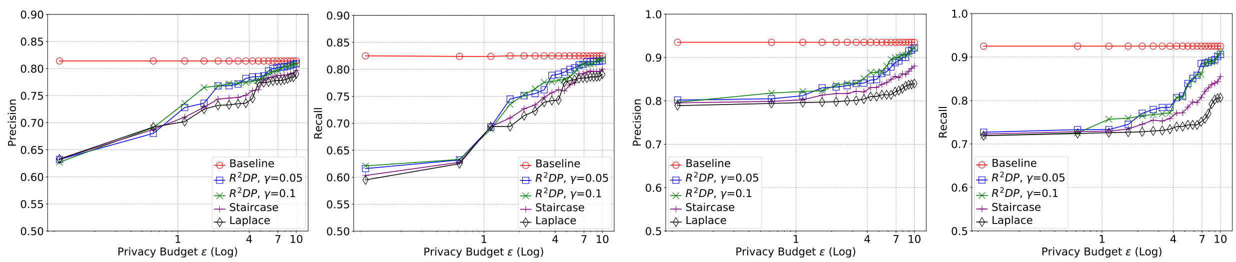
Differential privacy [23] is a model for preserving privacy while releasing the results of various useful functions, such as contingency tables, histograms and means [20]. Many existing works focus on improving the utility based on different mechanisms.

Noise Perturbation. Based on the general utility maximization framework from Ghosh et al. [35], Gupte and Sundararajan [38] further study the optimal noise probability distributions for single count queries. Later, Geng el al. [32, 33] demonstrate the optimal noise distribution has a Staircase-shaped PDF for Laplace mechanism. Furthermore, Balle and Wang [4] develop an optimal Gaussian mechanism in high privacy regime to minimize the noise and increase the utility for queries. Geng et al. [31] further show the optimal noise distribution is a uniform distribution over Gaussian mechanism. Moreover, Hardt et al. [41] study the privacy-utility trade-off for answering a set of linear queries over a histogram, where the error is defined as the worst expectation of the ℓ_2-norm (identical to variance) of the noise among all possible outputs. Subsequently, Brenner et al. [10] show that, for general query functions, no universally optimal DP mechanisms exist.



(a) $p = 1, \Delta q = 0.5$ (b) $p = 1, \Delta q = 1$ (c) $p = 2, \Delta q = 0.5$ (d) $p = 2, \Delta q = 1$

Figure 8: Mallows metric: R^{2}DP compared to Laplace and Staircase mechanisms for degree distribution (Facebook dataset).



(a) Precision vs. ϵ (UCI Adult) (b) Recall vs. ϵ (UCI Adult) (c) Precision vs. ϵ (KDDCUP99) (d) Recall vs. ϵ (KDDCUP99)

Figure 9: Accuracy evaluation for classification (UCI Adult dataset and KDDCUP99 dataset)

Sampling and Aggregation. Sampling and aggregation frameworks mostly split the database into chunks, and aggregate the result using a DP algorithm after querying each chunk [64]. To expand the applicability of output perturbation, Nissim et al. [64] propose a framework to formally analyze the effect of instance-based noise. Observing the highly compressible nature of many real-life data, researchers propose lossy compression techniques to add noise calibrated to the compressed data. Acs et al. [3] propose an optimization of Fourier perturbation algorithm that clusters and exploits the redundancy between bins. Instead of directly adding noise to histogram counts, it first lossily compresses the data, then adds noise calibrated to the data. Li et al. [56] propose an algorithm to partitions a data domain into uniform regions and adapts the strategy to fit the specific set of range queries to achieve a lower error rate. Zhang et al. [83] improve the clustering mechanism by sorting histogram bins based on the noisy counts.

Data Composition. Barak et al. [5] propose transforming the data into the Fouier domain, which could avoid the violation of consistency for low-order marginals in database tables. As efficiency is the main bottleneck for this approach when the number of attributes is large, Hay et al. [43] ensure that the error rate does not grow with the size of a database. The proposed hierarchical histogram method also achieves a lower error for a fixed domain. Different from one-dimensional datasets solution proposed by Hay et al. [43], Xiao et al. [79] propose *Privelet* that improves accuracy on datasets with arbitrary dimensions, which could reduce error to 25% compared to 70% as baseline error rate. Cormode et al. [18] apply *quadtrees* and *kd-trees* as new techniques for parameter setting to improve the accuracy on spatial data. Ding et al. [19] introduce a general noise-control framework on data cubes. Li et al. [58] unify the two range queries over histograms into one framework. Other techniques, such as principal component analysis (PCA), linear discriminant analysis (LDA) [48], and random projection [15, 80] are also used to lower the data dimension for reducing the errors. Cormode et al. [18] apply quadtrees (*data-independent*) and kd-trees (*data-dependent*) to add noise to a histogram output.

Adaptive Queries. In this technique, the improvement of utilities takes advantage of a known set of queries, for example, Dwork et al. [27] propose *Boosting for Queries* algorithm to obtain a better accuracy of learning algorithms. Hardt et al. [39, 40] present multiplicative weights mechanism to improve the efficiency of interactive queries. Instead of polynomial running time [25], this work achieves a nearly linear running time with a relaxed utility requirement. Yuan et al. [81, 82] propose low-rank mechanism (LRM) to further improve the adaptive queries. Other techniques such as correlated noise [63] and sparse vector technique (SVT) [60] are also used in adaptive queries.

Applications. Many researchers also work on improving the utility for different types of data, such as, the Fourier Perturbation Algorithm (FPA$_k$) [67] in time-series data (e.g., location traces, web history, and personal health), *kd-trees* on spatial data [18], and matrix-valued query [14].

Summary. Our R^{2}DP framework provides a complementary approach to those existing works by providing the opportunity of searching for the maximal utility along an extra dimension. This framework also enables data recipients to specify their utility requirements and the computed parameter could be incorporated into existing solutions to further improve utility.

7 CONCLUSION

This paper has proposed the R^{2}DP framework as a universal solution for optimizing a variety of utility metrics requested in different applications. It can automatically identify a distribution that yields near-optimal utility, and hence is more practical for emerging applications. Specifically, we have shown that a differentially private mechanism could be defined based on a random variable which is itself distributed according to some parameterized distributions. We have also shown that such a mechanism could explicitly take into account both the privacy requirements and the utility requirements specified by the data owner and data recipient, respectively. We have formally analyzed the privacy guarantee of R^{2}DP based on the well-known Laplace mechanism and formally proved the improvement of utility over the baseline Laplace mechanism. Furthermore, we discuss the potential of applying R2DP to advanced algorithms, e.g., workload queries. Finally, our experimental results based on six different utility metrics for statistical queries, machine learning and social network, as well as one privacy metric, have demonstrated that R2DP could significantly improve the utility of differentially private solutions for a wide range of applications.

REFERENCES

[1] [n.d.]. Trove page for Hardy, G. H. (Godfrey Harold) (1877-1947).

[2] Gregory A Aarons, Amy E Green, Lawrence A Palinkas, Shannon Self-Brown, Daniel J Whitaker, John R Lutzker, Jane F Silovsky, Debra B Hecht, and Mark J Chaffin. 2012. Dynamic adaptation process to implement an evidence-based child maltreatment intervention. *Implementation Science* 7, 1 (2012), 32.

[3] Gergely Acs, Claude Castelluccia, and Rui Chen. 2012. Differentially Private Histogram Publishing through Lossy Compression. In *12th IEEE International Conference on Data Mining (ICDM '12)*. Brussels, Belgium, 1–10.

[4] Borja Balle and Yu-Xiang Wang. 2018. Improving the Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising. In *Proceedings of the 35th International Conference on Machine Learning (ICML '18)*. Stockholm, Sweden, 403–412.

[5] Boaz Barak, Kamalika Chaudhuri, Cynthia Dwork, Satyen Kale, Frank McSherry, and Kunal Talwar. 2007. Privacy, Accuracy, and Consistency Too: A Holistic Solution to Contingency Table Release. In *Proceedings of the 26th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '07)*. ACM, Beijing, China, 273–282.

[6] Dimitri P Bertsekas. 2014. *Constrained optimization and Lagrange multiplier methods*. Academic press.

[7] Avrim Blum, Katrina Ligett, and Aaron Roth. 2008. A Learning Theory Approach to Non-interactive Database Privacy. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC '08)*. ACM, New York, NY, USA, 609–618.

[8] Nicolás E Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2014. Optimal geo-indistinguishable mechanisms for location privacy. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*. 251–262.

[9] Hans Bremermann. 1965. Distributions, complex variables, and Fourier transforms. (1965).

[10] Hai Brenner and Kobbi Nissim. 2010. Impossibility of Differentially Private Universally Optimal Mechanisms. In *IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS '10)*. Las Vegas, Nevada, USA, 71–80.

[11] Joshua W. S. Brown, Olga Ohrimenko, and Roberto Tamassia. 2013. Haze: Privacy-preserving Real-time Traffic Statistics. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL '13)*. ACM, New York, NY, USA, 540–543.

[12] Michael George Bulmer. 1979. *Principles of statistics*. Courier Corporation.

[13] T.-H. Hubert Chan, Elaine Shi, and Dawn Song. 2011. Private and Continual Release of Statistics. *ACM Transactions Information System Security* 14, 3, Article 26 (2011), 26:1–26:24 pages.

[14] Thee Chanyaswad, Alex Dytso, H. Vincent Poor, and Prateek Mittal. 2018. MVG Mechanism: Differential Privacy under Matrix-Valued Query. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS '18)*. Toronto, ON, Canada, 230–246.

[15] Thee Chanyaswad, Changchang Liu, and Prateek Mittal. 2019. RON-Gauss: Enhancing Utility in Non-Interactive Private Data Release. *PoPETs* 2019, 1 (2019), 26–46.

[16] Charalambos A. Charalambides. 2005. *Combinatorial Methods in Discrete Distributions (Wiley Series in Probability and Statistics)*. Vol. 600. Wiley-Interscience, New York, NY, USA.

[17] Joel E Cohen, YVES Derriennic, and GH Zbaganu. 1993. Majorization, monotonicity of relative entropy, and stochastic matrices. *Contemp. Math.* 149 (1993), 251–251.

[18] Graham Cormode, Cecilia M. Procopiuc, Divesh Srivastava, Entong Shen, and Ting Yu. 2012. Differentially Private Spatial Decompositions. In *IEEE 28th International Conference on Data Engineering (ICDE '12)*. IEEE Computer Society, Washington, DC, USA, 20–31.

[19] Bolin Ding, Marianne Winslett, Jiawei Han, and Zhenhui Li. 2011. Differentially private data cubes: Optimizing noise sources and consistency. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '11)*. Athens, Greece, 217–228.

[20] Cynthia Dwork. 2008. Differential Privacy: A Survey of Results. In *Theory and Applications of Models of Computation*, Vol. 4978. Springer Berlin Heidelberg, Berlin, Heidelberg, 1–19.

[21] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *25th Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer Berlin Heidelberg, Berlin, Heidelberg, 486–503.

[22] Cynthia Dwork and Jing Lei. 2009. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 371–380.

[23] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography*. Springer Berlin Heidelberg, Berlin, Heidelberg, 265–284.

[24] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. 2010. Differential Privacy Under Continual Observation. In *Proceedings of the 42rd ACM Symposium on Theory of Computing (STOC '10)*. ACM, New York, NY, USA, 715–724.

[25] Cynthia Dwork, Moni Naor, Omer Reingold, Guy N. Rothblum, and Salil P. Vadhan. 2009. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC '09)*. Bethesda, MD, USA, 381–390.

[26] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Found. Trends Theor. Comput. Sci.* 9, 3–4 (Aug. 2014), 211–407. https://doi.org/10.1561/0400000042

[27] Cynthia Dwork, Guy N. Rothblum, and Salil P. Vadhan. 2010. Boosting and Differential Privacy. In *Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science (FOCS '10)*. Las Vegas, Nevada, USA, 51–60.

[28] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the ACM SIGSAC conference on computer and communications security*. ACM, Scottsdale, AZ, USA, 1054–1067.

[29] Willliam Feller. 2008. *An introduction to probability theory and its applications*. Vol. 2. John Wiley & Sons.

[30] Marek Fisz. 2018. *Probability theory and mathematical statistics*. Vol. 3.

[31] Quan Geng, Wei Ding, Ruiqi Guo, and Sanjiv Kumar. 2018. Optimal Noise-Adding Mechanism in Additive Differential Privacy. *CoRR* abs/1809.10224 (2018). arXiv:1809.10224 http://arxiv.org/abs/1809.10224

[32] Quan Geng, Peter Kairouz, Sewoong Oh, and Pramod Viswanath. 2015. The Staircase Mechanism in Differential Privacy. *IEEE Journal of Selected Topics Signal Processing* 9, 7 (2015), 1176–1184.

[33] Quan Geng and Pramod Viswanath. 2014. The optimal mechanism in differential privacy. In *2014 IEEE International Symposium on Information Theory*. Honolulu, HI, USA, 2371–2375.

[34] Quan Geng and Pramod Viswanath. 2016. Optimal Noise Adding Mechanisms for Approximate Differential Privacy. *IEEE Transactions on Information Theory* 62, 2 (Feb 2016), 952–969.

[35] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. 2009. Universally Utility-maximizing Privacy Mechanisms. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC '09)*. ACM, New York, NY, USA, 351–360.

[36] Manuel Gil. 2011. *On Rényi divergence measures for continuous alphabet sources*. Ph.D. Dissertation. Citeseer.

[37] Yves Grandvalet and Yoshua Bengio. 2005. Semi-supervised learning by entropy minimization. In *Advances in neural information processing systems*. 529–536.

[38] Mangesh Gupte and Mukund Sundararajan. 2010. Universally Optimal Privacy Mechanisms for Minimax Agents. In *Proceedings of the 29th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '10)*. ACM, New York, NY, USA, 135–146.

[39] Moritz Hardt, Katrina Ligett, and Frank McSherry. 2012. A Simple and Practical Algorithm for Differentially Private Data Release. In *Proceedings of the 26th Annual Conference on Neural Information Processing Systems (NIPS '12)*. Lake Tahoe,Nevada, USA, 2348–2356.

[40] Moritz Hardt and Guy N. Rothblum. 2010. A Multiplicative Weights Mechanism for Privacy-Preserving Data Analysis. In *Proceedings of the 51th Annual IEEE Symposium on Foundations of Computer Science (FOCS '10)*. Las Vegas, Nevada, USA, 61–70.

[41] Moritz Hardt and Kunal Talwar. 2010. On the Geometry of Differential Privacy. In *Proceedings of the 42rd ACM Symposium on Theory of Computing (STOC '10)*. ACM, New York, NY, USA, 705–714.

[42] Michael Hay, Chao Li, Gerome Miklau, and David Jensen. 2009. Accurate Estimation of the Degree Distribution of Private Networks. In *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining (ICDM '09)*. IEEE Computer Society, Washington, DC, USA, 169–178. https://doi.org/10.1109/ICDM.2009.11

[43] Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. 2010. Boosting the Accuracy of Differentially Private Histograms Through Consistency. *VLDB* 3, 1-2 (Sept. 2010), 1021–1032.

[44] Ali Inan, Murat Kantarcioglu, Gabriel Ghinita, and Elisa Bertino. 2010. Private record matching using differential privacy. In *Proceedings of the 13th International Conference on Extending Database Technology*. 123–134.

[45] MV Jambunathan. 1954. Some properties of beta and gamma distributions. *The annals of mathematical statistics* 25, 2 (1954), 401–405.

[46] Johan Ludwig William Valdemar Jensen. 1906. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica* 30, 1 (1906), 175–193.

[47] Zhanglong Ji, Zachary C Lipton, and Charles Elkan. 2014. Differential privacy and machine learning: a survey and review. *arXiv preprint arXiv:1412.7584* (2014).

[48] Xiaoqian Jiang, Zhanglong Ji, Shuang Wang, Noman Mohammed, Samuel Cheng, and Lucila Ohno-Machado. 2013. Differential-Private Data Publishing Through Component Analysis. *Transactions Data Privacy* 6, 1 (2013), 19–34.

[49] Shiva Prasad Kasiviswanathan, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2013. Analyzing Graphs with Node Differential Privacy. In *Theory of Cryptography*, Amit Sahai (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 457–476.

[50] R. Bruce Kellogg. 1996. CRC standard mathematical tables and formulae. *SIAM Rev.* 38, 4 (1996), 691–692.

[51] Ron Kohavi. 1996. Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-tree Hybrid. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD' 96)*. AAAI Press, 202–207.

[52] Fragkiskos Koufogiannis, Shuo Han, and George J Pappas. 2015. Optimality of the laplace mechanism in differential privacy. *arXiv preprint arXiv:1504.00065* (2015).

[53] Jerome Le Ny and George J Pappas. 2014. Differentially private filtering. *IEEE Trans. Automat. Control* 59, 2 (2014), 341–354.

[54] Sangmin Lee, Edmund L Wong, Deepak Goel, Mike Dahlin, and Vitaly Shmatikov. 2013. πbox: A platform for privacy-preserving apps. In *Proceedings of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} '13)*. Lombard, IL, USA, 501–514.

[55] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[56] Chao Li, Michael Hay, Gerome Miklau, and Yue Wang. 2014. A Data- and Workload-aware Algorithm for Range Queries Under Differential Privacy. *VLDB* 7, 5 (Jan. 2014), 341–352.

[57] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor. 2010. Optimizing linear counting queries under differential privacy. In *Proceedings of the twenty-ninth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 123–134.

[58] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor. 2010. Optimizing linear counting queries under differential privacy. In *Proceedings of the 29th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '10)*. ACM, Indianapolis, Indiana, USA, 123–134.

[59] Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. 2015. The matrix mechanism: optimizing linear counting queries under differential privacy. *VLDB J.* 24, 6 (2015), 757–781. https://doi.org/10.1007/s00778-015-0398-x

[60] Min Lyu, Dong Su, and Ninghui Li. 2017. Understanding the Sparse Vector Technique for Differential Privacy. *PVLDB* 10, 6 (2017), 637–648.

[61] Frank McSherry. 2009. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '09)*. ACM, Rhode Island, USA, 19–30.

[62] Ilya Mironov. 2017. Rényi differential privacy. In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. IEEE, 263–275.

[63] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. 2013. The geometry of differential privacy: the sparse and approximate cases. In *Symposium on Theory of Computing Conference (STOC '13)*. Palo Alto, CA, USA, 351–360.

[64] Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. 2007. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC '07)*. San Diego, California, USA, 75–84.

[65] Jerome Le Ny and Meisam Mohammady. 2018. Differentially Private MIMO Filtering for Event Streams. *IEEE Trans. Automat. Control* 63, 1 (Jan 2018), 145–157.

[66] Harry H. Panjer. 1981. Recursive Evaluation of a Family of Compound Distributions. *ASTIN Bulletin* 12, 1 (1981), 22–26. https://doi.org/10.1017/S0515036100006796

[67] Vibhor Rastogi and Suman Nath. 2010. Differentially private aggregation of distributed time-series with transformation and encryption. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '10)*. Indianapolis, Indiana, USA, 735–746.

[68] Alfréd Rényi et al. 1961. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.

[69] Stefan Rolweicz and E Bedarczuk. 1986. *Functional Analysis and Control Theory: Linear Systems*. Kluwer Academic Publishers, USA.

[70] S. J. Stolfo, Wei Fan, Wenke Lee, A. Prodromidis, and P. K. Chan. 2000. Cost-based modeling for fraud and intrusion detection: results from the JAM project. In *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00*, Vol. 2. 130–144 vol.2. https://doi.org/10.1109/DISCEX.2000.821515

[71] Jaideep Vaidya, Basit Shafiq, Anirban Basu, and Yuan Hong. 2013. Differentially Private Naive Bayes Classification. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 01 (WI-IAT '13)*. IEEE Computer Society, USA, 571–576. https://doi.org/10.1109/WI-IAT.2013.80

[72] Jaideep Vaidya, Basit Shafiq, Anirban Basu, and Yuan Hong. 2013. Differentially Private Naive Bayes Classification. In *2013 IEEE/WIC/ACM International Conferences on Web Intelligence (WI '13)*. Atlanta, GA, USA, 571–576.

[73] Tim Van Erven and Peter Harremos. 2014. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* 60, 7 (2014), 3797–3820.

[74] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud. 2014. Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems. In *53rd IEEE Conference on Decision and Control*. 2130–2135. https://doi.org/10.1109/CDC.2014.7039713

[75] Y. Wang, Z. Huang, S. Mitra, and G. E. Dullerud. 2014. Entropy-minimizing mechanism for differential privacy of discrete-time linear feedback systems. In *53rd IEEE Conference on Decision and Control*. 2130–2135. https://doi.org/10.1109/CDC.2014.7039713

[76] Yue Wang, Xintao Wu, and Donghui Hu. 2016. Using Randomized Response for Differential Privacy Preserving Data Collection. In *Proceedings of EDBT/ICDT Workshops Joint Conference (EDBT/ICDT '16)*. Bordeaux, France.

[77] Christopher R. Wren, Yuri A. Ivanov, Darren Leigh, and Jonathan Westhues. 2007. The MERL Motion Detector Dataset. In *Proceedings of the 2007 Workshop on Massive Datasets (MD '07)*. ACM, New York, NY, USA, 10–14. https://doi.org/10.1145/1352922.1352926

[78] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. 2010. Differential privacy via wavelet transforms. *IEEE Transactions on knowledge and data engineering* 23, 8 (2010), 1200–1214.

[79] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. 2011. Differential Privacy via Wavelet Transforms. *IEEE Transactions on Knowledge and Data Engineering* 23, 8 (Aug. 2011), 1200–1214.

[80] Chugui Xu, Ju Ren, Yaoxue Zhang, Zhan Qin, and Kui Ren. 2017. DPPro: Differentially Private High-Dimensional Data Release via Random Projection. *IEEE Transactions Information Forensics and Security* 12, 12 (2017), 3081–3093.

[81] Ganzhao Yuan, Zhenjie Zhang, Marianne Winslett, Xiaokui Xiao, Yin Yang, and Zhifeng Hao. 2012. Low-Rank Mechanism: Optimizing Batch Queries under Differential Privacy. *PVLDB* 5, 11 (2012), 1352–1363.

[82] Ganzhao Yuan, Zhenjie Zhang, Marianne Winslett, Xiaokui Xiao, Yin Yang, and Zhifeng Hao. 2015. Optimizing Batch Linear Queries under Exact and Approximate Differential Privacy. *ACM Transactions Database Systems* 40, 2 (2015), 11:1–11:47.

[83] Xiaojian Zhang, Rui Chen, Jianliang Xu, Xiaofeng Meng, and Yingtao Xie. 2014. Towards Accurate Histogram Publication under Differential Privacy. In *Proceedings of the SIAM International Conference on Data Mining (SDM '14)*. Philadelphia, Pennsylvania, USA, 587–595.

[84] Zuhe Zhang, Benjamin IP Rubinstein, and Christos Dimitrakakis. 2016. On the differential privacy of Bayesian inference. In *Thirtieth AAAI Conference on Artificial Intelligence*.

APPENDIX

A DEMONSTRATION OF THEOREM 3.1

A Laplace distribution is of a ($\propto x \cdot e^{x \cdot t}$) order, where x is the inverse of the scale parameter. Second, since $x \cdot e^{x \cdot t} = \frac{\mathrm{d}e^{x \cdot t}}{\mathrm{d}t}$, the cumulative distribution function (CDF) resulted from randomizing x can be expressed in terms of the expectation $\mathbb{E}(e^{x \cdot t})$. We note that from now on, we will simply refer to R^{2}DP with Laplace distribution as the first fold PDF as *the R^{2}DP mechanism.*

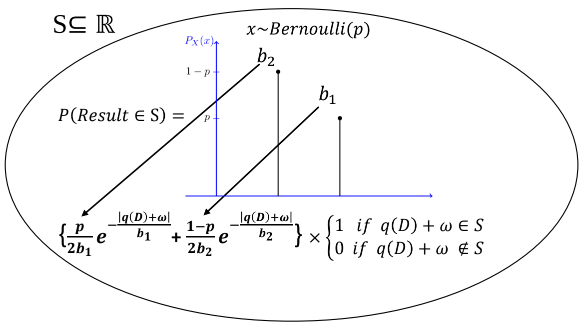


Figure 10: The term in the parenthesis is the derivative of $\mathbb{E}(e^{\frac{1}{b} \cdot -|w|})$ w.r.t. $-|w|$, and hence the above probability can be expressed in terms of the expectation

Example A.1. *Following Example 3.1, for a Bernoulli distributed scale parameter b, Figure 10 illustrates the above finding (see Appendix C for proof). It can be verified that the term inside the braces is the derivative of $\mathbb{E}(e^{\frac{1}{b} \cdot -|w|})$ w.r.t. $-|w|$, and hence the above probability can be expressed in terms of the expectation.*

B CASE STUDY PDFS

B.0.1 Discrete Probability Distributions. First, we consider two different mixture Laplace distributions that can be applied for constructing R^{2}DP with discrete probability distribution f_b.

(1) **Degenerate distribution.** A degenerate distribution is a probability distribution in a (discrete or continuous) space with support only in a space of lower dimension [9]. If the degenerate distribution is uni-variate (involving only a single random variable), it will be a deterministic distribution and takes only a single value. Therefore, the degenerate distribution is identical to the baseline Laplace mechanism as it also assigns the mechanism one single scale parameter b_0. Specifically, the probability mass function of the uni-variate degenerate distribution is:

$$f_{\delta,k_0}(x) = \begin{cases} 1 & x = k_0 \\ 0 & x \neq k_0 \end{cases}$$

The MGF for the degenerate distribution δ_{k_0} is given by $M_k(t) = e^{t \cdot k_0}$ [12]. Using Equation 4, Theorem B.1 gives the same DP guarantee as the baseline Laplace mechanism.

Theorem B.1. *The R^{2}DP mechanism $M_q(d, \epsilon)$, $\epsilon \sim f_{\delta, \frac{1}{b_0}}(\epsilon)$, is $\frac{\Delta q}{b_0}$-differentially private.*

Obviously, this distribution does not improve the bound in Theorem 4.5 but shows the soundness of our findings.

(2) **Bernoulli distribution.** The probability mass function of this distribution, over possible outcomes k, is

$$f_B(k; p) = \begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0. \end{cases}$$

Note that the binary outcomes $k = 0$ and $k = 1$ can be mapped to any two outcomes X_0 and X_1, respectively. Therefore, we consider the following Bernoulli outcomes

$$f_{B,X_0,X_1}(X; p) = \begin{cases} p & \text{if } X = X_1, \\ 1-p & \text{if } X = X_0. \end{cases}$$

The MGF for Bernoulli distribution $f_{B,X_0,X_1}(X; p)$ is $M_X(t) = p \cdot e^{t \cdot X_0} + (1-p) \cdot e^{t \cdot X_1}$ [12]. We now derive the precise differential privacy guarantee of an R^{2}DP mechanism with its scale parameter randomized according to a Bernoulli distribution.

Theorem B.2. *The R^{2}DP mechanism $M_q(d, \epsilon)$, $\epsilon \sim f_{B, \frac{1}{b_0}, \frac{1}{b_1}}(\epsilon; p)$, satisfies $\ln[p \cdot e^{\frac{\Delta q}{b_0}} + (1-p) \cdot e^{\frac{\Delta q}{b_1}}]$ differential privacy.*

This bound is exactly the mean of $e^{\epsilon(b)}$ given in Theorem 4.5.

B.0.2 Continuous Probability Distributions. We now investigate three compound Laplace distributions.

(1) **Gamma distribution.** The gamma distribution is a two-parameter family of continuous probability distributions with a shape parameter $k > 0$ and a scale parameter θ. Besides the generality, the gamma distribution is the maximum entropy probability distribution (both w.r.t. a uniform base measure and w.r.t. a $1/x$ base measure) for a random variable X for which $\mathbb{E}(X) = k\theta = \alpha/\beta$ is fixed and greater than zero, and $\mathbb{E}[\ln(X)] = \psi(k) + \ln(\theta) = \psi(\alpha) - \ln(\beta)$ is fixed (ψ is the digamma function). Therefore, it may provide a relatively higher privacy-utility trade-off in comparison to the other candidates [45, 50]. A random variable X that is gamma-distributed with shape α and rate β is denoted by $X \sim \Gamma(k, \theta)$ and the corresponding PDF is

$$f_\Gamma(X; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k) \cdot \theta^k} \quad \text{for } X > 0 \text{ and } k, \theta > 0,$$

where $\Gamma(\alpha)$ is the gamma function. We now investigate the differential privacy guarantee provided by assuming that the reciprocal of the scale parameter b in Laplace mechanism is distributed according to the gamma distribution (see Appendix C for the proof).

Theorem B.3. *The R^{2}DP mechanism $M_q(d, \epsilon)$, $\epsilon \sim f_\Gamma(\epsilon; k, \theta)$, satisfies $\big((k+1) \cdot \ln(1 + \Delta q \cdot \theta)\big)$ differential privacy.*

We now apply the necessary condition given in Equation 8 (see Appendix C for the proof).

Lemma B.4. *R^{2}DP using Gamma distribution can satisfy the necessary condition in Equation 8.*

Therefore, Gamma distribution may improve over the baseline, and this can be computed by optimizing the privacy-utility trade-off using the Lagrange multiplier function in Equation 6. Also, our numerical results show that, this distribution is more effective for large ϵ (weaker privacy guarantees).

(2) **Uniform distribution.** In probability theory and statistics, the continuous uniform distribution or rectangular distribution is a family of symmetric probability distributions such that for each member of the family, all intervals of the same length on the support of the distribution are equally probable. The support is defined by the two parameters, a and b, which are the minimum and maximum values. The distribution is often abbreviated as $U(a,b)$, which is the maximum entropy probability distribution for a random variable X under no constraint; other than that, it is contained in the distribution's support [45, 50]. The MGF for $U(a,b)$ is

$$M_X(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & \text{for } t \neq 0, \\ 1 & \text{for for } t = 0. \end{cases}$$

Using Theorem 4.1, we now drive the precise differential privacy guarantee of an R^{2}DP mechanism for uniform distribution $U(a,b)$.

Theorem B.5. *The R^{2}DP mechanism $M_q(d,\epsilon)$, $\epsilon \sim f_{U(a,b)}(\epsilon)$, is $\ln\left[\frac{\alpha^2-\beta^2}{2((1+\beta)e^{-\beta}-(1+\alpha)e^{-\alpha})}\right]$-differentially private, where $\alpha = a\cdot\Delta q$ and $\beta = b\cdot\Delta q$.*

We now apply the necessary condition given in Equation 8. One can easily verify that the inequality holds for an infinite number of settings, e.g., $a = 0.5$, $b = 9$ and $\Delta q = 1.2$.

Lemma B.6. *R^{2}DP using uniform distribution can satisfy the necessary condition in Equation 8.*

Therefore, R^{2}DP using uniform distribution may improve over the baseline, and this can be computed by optimizing the privacy-utility trade-off using the Lagrange multiplier function in Equation 6. Also, our numerical results show that, this distribution can also be effective for both small and large ϵ.

(3) **Truncated Gaussian distribution.** The last distribution we consider is the Truncated Gaussian distribution. This distribution is derived from that of a normally distributed random variable by bounding the random variable from either below or above (or both). Therefore, we can benefit from the numerous useful properties of Gaussian distribution, by truncating the negative region of the Gaussian distribution. Suppose $X \sim \mathcal{N}(\mu,\sigma^2)$ has a Gaussian distribution and lies within the interval $X \in (a,b)$, $-\infty \leq a < b \leq \infty$. Then, X conditional on $a < X < b$ has a truncated Gaussian distribution with the following probability density function

$$f_{\mathcal{N}^T}(X;\mu,\sigma,a,b) = \frac{\phi(\frac{X-\mu}{\sigma})}{\sigma\cdot\left(\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma})\right)} \quad \text{for } a \leq x \leq b$$

and by $f_{\mathcal{N}^T} = 0$ otherwise. Here, $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ and $\Phi(x) = 1-Q(x)$ are PDF and CDF of the standard Gaussian distribution, respectively. Next, using Theorem 4.1, we give the differential privacy guarantee provided by the mechanism assuming that the reciprocal of b is distributed according to the truncated Gaussian distribution.

Theorem B.7. *The R^{2}DP mechanism $\mathcal{M}_q(d,\epsilon)$, $\epsilon \sim f_{\mathcal{N}^T}(\epsilon;\mu,\sigma,a,b)$, satisfies $\epsilon_{\mathcal{N}^T}$- differential privacy, where*

$$\epsilon_{\mathcal{N}^T} = \ln\left[\frac{\mu + \dfrac{\sigma\cdot(\phi(\alpha)-\phi(\beta))}{(\Phi(\beta)-\Phi(\alpha))}}{\dfrac{\mathrm{d}M_{\mathcal{N}^T}(t)}{\mathrm{d}t}|_{t}=-\Delta q}\right] \tag{9}$$

in which $\phi(\cdot)$ is the probability density function of the standard normal distribution, $\phi(\cdot)$ is its cumulative distribution function and $\alpha = \frac{a-\mu}{\sigma}$ and $\beta = \frac{b-\mu}{\sigma}$.

Lemma B.8 (see Appendix C for the proof). *R^{2}DP using truncated Gaussian distribution can satisfy the necessary condition in Equation 8.*

Therefore, truncated Gaussian distribution may improve over the baseline, and this can be computed by optimizing the privacy-utility trade-off using the Lagrange multiplier function in Equation 6. In particular, our numerical results show that, this distribution can also be effective for smaller ϵ (stronger privacy guarantees).

C PROOFS

Example A.1. Following Example 3.1, for a Bernoulli distributed scale parameter b, we have

$$\mathbb{P}(\mathcal{M}_q(d,b)\in S)$$
$$= \int_{\mathbb{R}} \frac{p}{2b_1}\cdot \mathbb{1}_S\{q(d)+w\}e^{\frac{-|w|}{b_1}} + \frac{1-p}{2b_2}\cdot \mathbb{1}_S\{q(d)+w\}e^{\frac{-|w|}{b_2}}\,dw$$
$$= \int_{\mathbb{R}} \Big(\frac{p}{2b_1}\cdot e^{\frac{-|w|}{b_1}} + \frac{1-p}{2b_2}\cdot e^{\frac{-|w|}{b_2}}\Big)\mathbb{1}_S\{q(d)+w\}dw$$

where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. It can be verified that the term in the braces is the derivative of $\mathbb{E}(e^{\frac{1}{b}\cdot -|w|})$ w.r.t. $-|w|$, and hence the above probability can be expressed in terms of the expectation. □

Theorem 3.1. For an R^{2}DP Laplace mechanism and $\forall S \subset \mathbb{R}$ measurable and dataset d in D, we have

$$\mathbb{P}(\mathcal{M}_q(d,b)\in S)$$
$$= \int_{\mathbb{R}_{\geq 0}} f(b)\frac{1}{2b}\int_{\mathbb{R}} \mathbb{1}_S\{q(d)+w\}e^{\frac{-|w|}{b}}\,dw\,db$$
$$= \int_{\mathbb{R}_{\geq 0}} g(u)\frac{u}{2}\int_{\mathbb{R}} \mathbb{1}_S\{q(d)+w\}e^{-|w|\cdot u}\,dw\,du$$
$$= \int_{\mathbb{R}} \mathbb{1}_S\{q(d)+w\}\int_{\mathbb{R}_{\geq 0}} g(u)\frac{u}{2}e^{-|w|\cdot u}\,du\,dw$$
$$= \int_{\mathbb{R}} \mathbb{1}_S\{q(d)+w\}\frac{1}{2}\frac{dM_u(t)}{dt}|_{t=-|w|}\,dw$$
$$= \frac{1}{2}\int_S \frac{dM_u(t)}{dt}|_{t=-|x-q(d)|}dx \tag{10}$$

$$= \frac{1}{2}\cdot\Big[-M_u(-|x-q(d)|)|_{S_{\geq q(d)}} + M_u(-|x-q(d)|)|_{S_{<q(d)}}\Big] \tag{11}$$

where $u = b^{-1}$, is reciprocal of random variable b and $g(u) = \frac{1}{u^2}\cdot f(\frac{1}{u})$. Note that $M_u(t)$ is the MGF of random variable u which is identical with $M_{\frac{1}{b}}(t)$. □

Theorem 4.1. To prove this theorem, we first need to give two lemmas on the properties of R^{2}DP Laplace mechanism and MGFs.

Lemma C.1. *The R^{2}DP mechanism $\mathcal{M}_q(d,b)$, is*

$$\ln\left[\max_{\forall x\in\mathbb{R}}\left\{\frac{\dfrac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d)|}}{\dfrac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d')|}}\right\}\right]\text{-differentially private.} \tag{12}$$

Proof. According to Equation 10,

$$\mathbb{P}(\mathcal{M}_q(d,b) \in S) = \frac{1}{2}\int_S \frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d)|}dx$$

$$= \frac{1}{2}\int_S \frac{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d)|}}{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d')|}} \cdot \frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d')|}dx$$

Denote by

$$e^{\epsilon} = \sup\left\{\frac{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d)|}}{\frac{\mathrm{d}M_{\frac{1}{b}}(t)}{\mathrm{d}t}|_{t=-|x-q(d')|}}, \forall x \in S\right\},$$

$$\Rightarrow \mathbb{P}(\mathcal{M}_q(d,b) \in S) \leq e^{\epsilon} \cdot \mathbb{P}(\mathcal{M}_q(d',b) \in S)$$

and the choice of $S = \mathbb{R}$ concludes the proof. □

Next, we show the log-convexity property of the first derivative of moment generating functions.

Lemma C.2. *First derivative of a moment generating function defined by* $\frac{\mathrm{d}M(t)}{\mathrm{d}t} = \mathbb{E}(z \cdot e^{zt})$ *is log-convex.*

Proof. For real- or complex-valued random variables X and Y, Hölder's inequality [1] reads; $\mathbb{E}(|XY|) \leq (\mathbb{E}(|X|)^p)^{1/p} \cdot (\mathbb{E}(|Y|)^q)^{1/q}$ for any $1 < p, q < \infty$ with $1/p + 1/q = 1$. Next, for all $\theta \in (0,1)$ and $0 \leq x_1, x_2 < \infty$, define $X = z^{\theta} \cdot e^{\theta x_1 z}$, $Y = z^{1-\theta} \cdot e^{(1-\theta)x_2 z}$ and $p = 1/\theta$, $q = 1/(1-\theta)$. Therefore, we have

$$\mathbb{E}(z \cdot e^{(\theta x_1 + (1-\theta)x_2 z)}) \leq \mathbb{E}(z \cdot e^{x_1 z})^{\theta} \cdot \mathbb{E}(z \cdot e^{x_2 z})^{1-\theta}$$

which shows the definition of log-convexity holds for $M'(t)$. □

Back to the original proof, following the DP guarantee in Lemma C.1, and using triangle inequality, we have

$$e^{\epsilon} = \max_{\forall x \in \mathbb{R}}\left\{\frac{\mathbb{E}(\epsilon \cdot e^{(-|x-q(d)| \cdot \epsilon)})}{\mathbb{E}(\epsilon \cdot e^{(-|x-q(d')| \cdot \epsilon)})}\right\} \leq \max_{\forall t \in \mathbb{R}_{\leq 0}}\left\{\frac{\mathbb{E}(\epsilon \cdot e^{(t \cdot \epsilon)})}{\mathbb{E}(\epsilon \cdot e^{((t-\Delta q) \cdot \epsilon)})}\right\}$$

Next, we show that $f(t) = \frac{\mathbb{E}(\epsilon \cdot e^{(t \cdot \epsilon)})}{\mathbb{E}(\epsilon \cdot e^{((t-\Delta q) \cdot \epsilon)})}$ is non-decreasing w.r.t. t. For this purpose, we must show that

$$f'(t) = \frac{M''(t) \cdot M'(t-\Delta q) - M'(t) \cdot M''(t-\Delta q)}{M'^2(t-\Delta q)}$$

is non-negative. However, this is equivalent to show that $\frac{M''(t)}{M'(t)} \geq \frac{M''(t-\Delta q)}{M'(t-\Delta q)}$ or more generally $\frac{M''(t)}{M'(t)}$ is not-decreasing. However, following the log-convexity of first $M'(t)$, the logarithmic derivative of $M'(t)$ denoted by $\frac{M''(t)}{M'(t)}$ is non-decreasing. Thus, for all $t < 0$, $f(t) \leq f(0)$, and evaluating $e^{\epsilon(t)}$ at $t = 0$, concludes our proof. □

Theorem 4.5. Following Theorem 2.2, an ϵ-DP Laplace mechanism is $(\gamma, e^{\frac{-\gamma}{b(\epsilon)}})$-useful for all $\gamma \geq 0$, where $b(\epsilon) = \frac{\Delta q}{\epsilon}$. Therefore, for the usefulness of the baseline Laplace mechanism at $\epsilon = \ln[\mathbb{E}_{\frac{1}{b}}(e^{\epsilon(b)})]$, we have

$$e^{\frac{-\gamma \cdot \ln[\mathbb{E}_{\frac{1}{b}}(e^{\epsilon(b)})]}{\Delta q}} = \left(\mathbb{E}_{\frac{1}{b}}(e^{\epsilon(b)})\right)^{\frac{-\gamma}{\Delta q}} = \left(\mathbb{E}_{\frac{1}{b}}(e^{\frac{\Delta q}{b}})\right)^{\frac{-\gamma}{\Delta q}} \leq \mathbb{E}_{\frac{1}{b}}\left(e^{\frac{-\gamma}{b}}\right)$$

where the last inequality relation is verified by Jensen inequality [46] as $g(x) = x^{\frac{-\gamma}{b}}$ is a convex function. Recall the following Jensen inequality: Let $(\Omega, \mathfrak{F}, \mathrm{P})$ be a probability space, X an integrable real-valued random variable and g a convex function. Then

$$g(\mathbb{E}(X)) \leq \mathbb{E}(g(X))$$

Therefore,

$$1 - e^{\frac{-\gamma \cdot \ln[\mathbb{E}_{\frac{1}{b}}(e^{\epsilon(b)})]}{\Delta q}} \geq 1 - \mathbb{E}_{\frac{1}{b}}\left(e^{\frac{-\gamma}{b}}\right) = U(\ln[\mathbb{E}_{\frac{1}{b}}(e^{\epsilon(b)})], \Delta q, \gamma)$$

This completes the proof. □

Theorem B.1. For $\frac{1}{b} \sim f_{\delta, \frac{1}{b_0}}(\frac{1}{b})$, the MGF is given by $M_{\frac{1}{b}}(t) = e^{\frac{t}{b_0}}$. Following Theorem C.1, one can write

$$e^{\epsilon} = \max_{\forall x \in \mathbb{R}}\left\{\frac{\frac{1}{b_0} \cdot e^{\frac{-|x-q(d)|}{b_0}}}{\frac{1}{b_0} \cdot e^{\frac{-|x-q(d')|}{b_0}}}\right\} = \max_{\forall x \in \mathbb{R}}\left\{e^{\frac{|x-q(d')|-|x-q(d)|}{b_0}}\right\}$$

$$\leq \max_{\forall x \in \mathbb{R}}\left\{e^{\frac{|q(d)-q(d')|}{b_0}}\right\} = e^{\frac{\Delta q}{b_0}}$$

where the last inequality is from triangle inequality. □

Theorem B.2. The R^{2}DP Laplace mechanism $\mathcal{M}_q(d,b)$, $\frac{1}{b} \sim f_{B, \frac{1}{b_0}, \frac{1}{b_1}}(\frac{1}{b}; p)$ returns with probability p, a Laplace mechanism with scale parameter b_1, and with probability $1-p$ another Laplace mechanism with scale parameter b_2. To this end, we are looking for

$$e^{\epsilon} = \max_{\forall x \in \mathbb{R}}\left\{\frac{\frac{p}{b_0} \cdot e^{\frac{-|x-q(d)|}{b_0}} + \frac{1-p}{b_1} \cdot e^{\frac{-|x-q(d)|}{b_1}}}{\frac{p}{b_0} \cdot e^{\frac{-|x-q(d')|}{b_0}} + \frac{1-p}{b_1} \cdot e^{\frac{-|x-q(d')|}{b_1}}}\right\}$$

Therefore, using triangle inequality, we have

$$e^{\epsilon_1} = \max_{\forall S \in \mathbb{R}}\left\{\frac{p \cdot e^{\frac{-|x-q(d)|}{b_0}} + (1-p) \cdot e^{\frac{-|x-q(d)|}{b_1}}}{p \cdot e^{\frac{-|x-q(d')|}{b_0}} + (1-p) \cdot e^{\frac{-|x-q(d')|}{b_1}}}\right\}$$

$$\leq \max_{\forall x \geq q(d)}\left\{\frac{p \cdot e^{\frac{\Delta q-|x-q(d')|}{b_0}} + (1-p) \cdot e^{\frac{\Delta q+-|x-q(d')|}{b_1}}}{p \cdot e^{\frac{-|x-q(d')|}{b_0}} + (1-p) \cdot e^{\frac{-|x-q(d')|}{b_1}}}\right\}$$

Let us make the substitutions $X = e^{\frac{-|x-q(d')|}{b_0}}$, $a = e^{\frac{\Delta q}{b_0}}$ and k$=\frac{b_0}{b_1} > 1$. Hence, we have

$$e^{\epsilon} \leq \max_{\forall X \in (0,1)}\left\{\frac{p \cdot a \cdot X + (1-p) \cdot (a \cdot X)^k}{p \cdot X + (1-p) \cdot X^k}\right\}$$

To obtain e^{ϵ}, we need to find all the critical points of $e^{\epsilon_1}(X) = \frac{p \cdot a \cdot X + (1-p) \cdot (a \cdot X)^k}{p \cdot X + (1-p) \cdot X^k}$. However, the critical points of a fractional function are the roots of the numerator of its derivative. Hence, suppose

$$\frac{\mathrm{d}e^{\epsilon}(X)}{\mathrm{d}X} = \frac{N(X)}{D(X)}$$

then

$$\Rightarrow N(X) = \left(p \cdot a + (1-p) \cdot k \cdot a \cdot (a \cdot X)^{k-1}\right)$$
$$\cdot\left(p \cdot X + (1-p) \cdot X^k\right) - \left(p + (1-p) \cdot k \cdot X^{k-1}\right)$$
$$\cdot\left(p \cdot a \cdot X + (1-p) \cdot (a \cdot X)^k\right)$$
$$= p \cdot (1-p) \cdot (k-1) \cdot (a^{k-1} - 1) \cdot X^k$$

However, all the terms in the last expression are strictly positive. Therefore, the only critical points are $X = 0$ and $X = 1$ and as the function is strictly increasing,

$$e^{\epsilon} \leq e^{\epsilon}(1) = p \cdot a + (1-p) \cdot (a)^k$$
$$= p \cdot e^{\frac{\Delta q}{b_0}} + (1-p) \cdot e^{\frac{\Delta q}{b_1}}$$

which is the bound in the Theorem. □

Theorem B.3. For a Gamma distribution with shape parameters k and scale parameters θ, the MGF at point t is given as $(1-\theta \cdot t)^{-k}$. Since $\frac{1}{b} \sim f_{\Gamma}(\frac{1}{b}; k, \theta)$, following Theorem C.1, one can write

$$e^{\epsilon} = \max_{\forall x \in \mathbb{R}} \left\{ \frac{k \cdot \theta \cdot (1+\theta \cdot |x-q(d)|)^{-k-1}}{k \cdot \theta \cdot (1+\theta \cdot |x-q(d')|)^{-k-1}} \right\}$$
$$\Rightarrow \epsilon = \max_{\forall x \in \mathbb{R}} \left\{ (k+1) \cdot \ln \left[\frac{(1+\theta \cdot |x-q(d')|)}{(1+\theta \cdot |x-q(d)|)} \right] \right\}$$

to find the maximum of the ln term, denote by $X = 1+\theta \cdot |x-q(d)|$. Moreover, since $|x-q(d')| \leq |x-q(d)| + \Delta q$, we have

$$\Rightarrow \epsilon \leq \max_{\forall X \geq 1} \left\{ \frac{X + \Delta q \cdot \theta}{X} \right\}$$

However, since

$$\forall X \geq 1, \ \frac{X + \Delta q \cdot \theta}{X}$$

is strictly decreasing, we have

$$\Rightarrow \epsilon = (k+1) \cdot \ln \left[1 + \theta \cdot \Delta q \right]$$

This completes the proof. □

Lemma B.4. We need to show that there exist k and θ such that $(k+1) \cdot \ln(1+\Delta q \cdot \theta) < -k \cdot \ln(1-\Delta q \cdot \theta)$, $\theta < \frac{1}{\Delta q}$. Given $\theta = \frac{1}{2\Delta q}$, we need to show that $\exists k, k \cdot \ln(2) > (k+1) \cdot \ln(1.5)$, which always holds for all $k > 1.4094$. □

Lemma B.8. Using exhaustive search, suppose $\mu = 0.5223$,$\sigma = 1.5454$, $a = 0.5223$ and for $\epsilon = 1.1703$ and $\Delta q = 0.6$, we will get $\ln(M_{\mathcal{N}^T}(\Delta q)) = 1.2417$. □

D LAGRANGE MULTIPLIER FUNCTION

The Lagrange Multiplier Function (all possible linear combinations of the Gamma, uniform and truncated Gaussian distributions) is:

$$\begin{aligned} &\mathcal{L}(a_1, a_2, a_3, k, \theta, a_u, b_u, \mu, \sigma, a_{\mathcal{N}^T}, b_{\mathcal{N}^T}, \Lambda) \\ &= M_{\Gamma(k,\theta)}(-a_1\gamma) \cdot M_{U(a_u,b_u)}(-a_2\gamma) \\ &\cdot M_{\mathcal{N}^T(\mu,\sigma,a_{\mathcal{N}^T},b_{\mathcal{N}^T})}(-a_3\gamma) + \Lambda \cdot (\ln \left[\frac{\mathsf{N}}{\mathsf{D}} \right] - \epsilon) \end{aligned} \tag{13}$$

where the numerator and the denominator N, D are

$$\mathsf{N} = (a_1 \cdot k \cdot \theta) + (a_2 \cdot \tfrac{a+b}{2}) + (a_3 \cdot (\mu + (\frac{\sigma \cdot \phi(\alpha) - \phi(\beta))}{(\Phi(\beta) - \Phi(\alpha))}))$$

$$\begin{aligned} \mathsf{D} = &\ a_1 \cdot M'_{\Gamma(k,\theta)}(-a_1 \cdot \Delta q) \cdot M_{U(a_u,b_u)}(-a_2 \cdot \Delta q) \\ &\cdot M_{\mathcal{N}^T(\mu,\sigma,a_{\mathcal{N}^T},b_{\mathcal{N}^T})}(-a_3 \cdot \Delta q) \\ &+ a_2 \cdot M_{\Gamma(k,\theta)}(-a_1 \cdot \Delta q) \cdot M'_{U(a_u,b_u)}(-a_2 \cdot \Delta q) \\ &\cdot M_{\mathcal{N}^T(\mu,\sigma,a_{\mathcal{N}^T},b_{\mathcal{N}^T})}(-a_3 \cdot \Delta q) \\ &+ a_3 \cdot M_{\Gamma(k,\theta)}(-a_1 \cdot \Delta q) \cdot M_{U(a_u,b_u)}(-a_2 \cdot \Delta q) \\ &\cdot M'_{\mathcal{N}^T(\mu,\sigma,a_{\mathcal{N}^T},b_{\mathcal{N}^T})}(-a_3 \cdot \Delta q) \end{aligned}$$

E NUMERICAL ANALYSIS

We also demonstrate the effectiveness of R^2DP through numerical results based on Algorithm 1 (the ensemble R^2DP algorithm). In particular, Figure 11 depicts the corresponding usefulness (the probability of the results to be within a pre-specified error bound) of the R^2DP, the Laplace and the Staircase mechanisms. Figure 11 clearly demonstrates the fact that the R^2DP mechanism can significantly improve both already considered to be competing mechanisms. In particular, we observe the power of the R^2DP mechanism in generating very high utility results, e.g., results with more than 0.8 probability fallen inside only $\gamma = 0.1$ error-bound, owing to automatically searching a large search space of PDFs.

F R^2DP AND OTHER DP MECHANISMS

In this section we briefly discuss the application of the R^2DP framework in two other well-known baseline DP mechanisms.

F.1 R^2DP Exponential Mechanism

The exponential mechanism was designed for situations in which we wish to choose the "best" response but adding noise directly to the computed quantity can completely destroy its value, such as setting a price in an auction, where the goal is to maximize revenue, and adding a small amount of positive noise to the optimal price (in order to protect the privacy of a bid) could dramatically reduce the resulting revenue [26]. The exponential mechanism is the natural building block for answering queries with arbitrary utilities (and arbitrary non-numeric range), while preserving differential privacy. Given some arbitrary range $\mathcal{R}$, the exponential mechanism is defined with respect to some utility function $u : \mathbb{N}^{|\mathcal{X}|} \times \mathcal{R} \rightarrow \mathbb{R}$, which maps database/output pairs to utility scores. Intuitively, for a fixed database x, the user prefers that the mechanism outputs some element of $\mathcal{R}$ with the maximum possible utility score. Note that when we talk about the sensitivity of the utility score $u : \mathbb{N}^{|\mathcal{X}|} \times \mathcal{R} \rightarrow \mathbb{R}$, we care only about the sensitivity of u with respect to its database argument; it can be arbitrarily sensitive in its range argument:

$$\Delta u \equiv \max_{r \in \mathcal{R}} \max_{x,y:\|x-y\| \leq 1} |u(x,r) - u(y,r)|.$$

The intuition behind the exponential mechanism is to output each possible $r \in \mathcal{R}$ with probability proportional to $\exp(\epsilon u(x,r)/\Delta u)$ and so the privacy loss is approximately:

$$\ln \left(\frac{exp(\epsilon u(x,r)/\Delta u)}{exp(\epsilon u(y,r)/\Delta u)} \right) = \epsilon [u(x,r) - u(y,r)/\Delta u] \leq \epsilon \tag{14}$$

The exponential mechanism is a canonical ϵ-DP mechanism, meaning that it describes a class of mechanisms that includes all

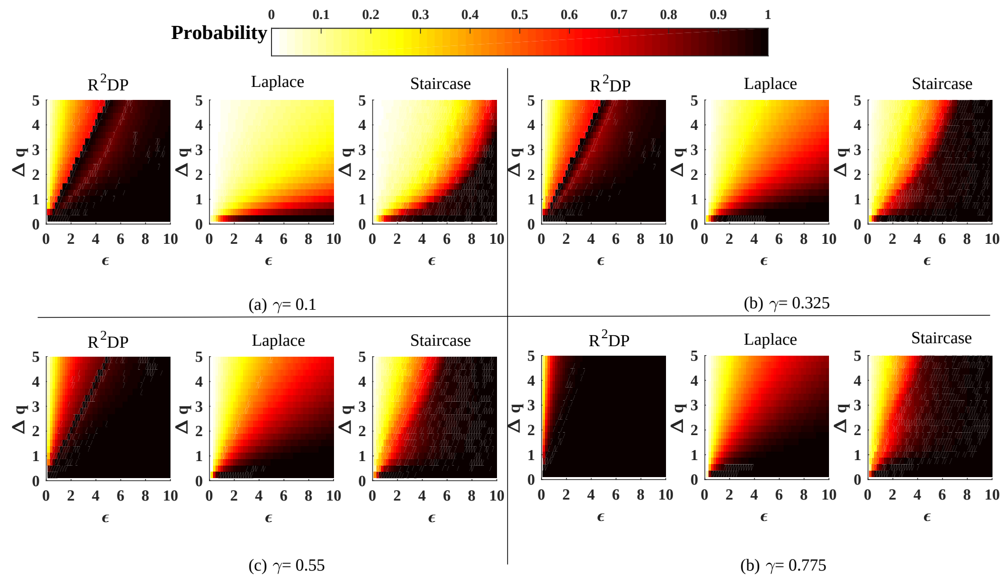


Figure 11: The R^2DP mechanism significantly outperforms the competing Laplace and the staircase mechanisms in maximizing the usefulness metric (an example of a utility metric with no known optimal PDF).

possible differentially private mechanisms. However, the exponential mechanism can define a complex distribution over a large arbitrary domain, and so it may not be possible to implement the exponential mechanism efficiently when the range of u is superpolynomially large in the natural parameters of the problem [26]. This is the main restrictive aspect of the exponential mechanism against leveraging different accuracy metrics. However, the exponential mechanism can benefit from the additional randomization of privacy budget, to handle the complexity (excessive sharpness) of the defined probability distribution. In particular, as we mentioned earlier, compound (or mixture) distributions arise naturally where a statistical population contains two or more sub-population which is the case for the exponential mechanism. Thus we motivate the application of the R^2DP framework in designing exponential mechanisms with rather smooth but accurate distributions around each element in the range of u. However, further discussion on R^2DP exponential mechanism requires formal analysis, e.g., deriving the DP guarantee of such a mechanism.

F.2 R^2DP and Differential Privacy Relaxations

R^2DP can also be studied under various relaxations of differential privacy, e.g., (ϵ, δ)-differential privacy or Rényi Differential Privacy [62] which is a privacy notion based on the Rényi divergence [73]. These relaxations allow suppressing the long tails of the mechanism's distribution where pure ϵ-differential privacy guarantees may not hold. Instead, they offer asymptotically smaller cumulative loss under composition and allow greater flexibility in the selection of privacy preserving mechanisms [62]. In the following, we briefly discuss the application of R^2DP in two of such relaxed notions of the differential privacy .

F.2.1 R^2DP Gaussian Mechanism. A relaxation of ϵ-differential privacy allows an additional bound δ in its defining inequality:

Definition F.1 ((ϵ, δ)-differential privacy [21]). *A randomized mechanism $M : \mathrm{D} \times \Omega \to \mathrm{R}$ is (ϵ, δ)-differentially private if for all adjacent $d, d' \in \mathrm{D}$, we have*

$$\mathbb{P}(M(d) \in S) \leq e^{\epsilon}\mathbb{P}(M(d') \in S) + \delta, \quad \forall S \subset \mathrm{R}. \tag{15}$$

This definition quantifies the allowed deviation (δ) for the output distribution of a ϵ-differentially private mechanism, when a single individual is added or removed from a dataset. A differentially private mechanism proposed in [21] modifies an answer to a numerical query by adding the independent and identically distributed zero-mean Gaussian noise.

Given the definition of the Q-function $Q(x) := \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du$, we have the following theorem [21, 53].

Theorem F.1. *Let $q : \mathrm{D} \to \mathbb{R}$ be a query and $\epsilon > 0$. Then the Laplace mechanism $\mathcal{M}_q : \mathrm{D} \times \Omega \to \mathbb{R}$ defined by $\mathcal{M}_q(d) = q(d) + w$, with $w \sim \mathcal{N}\left(0, \sigma^2\right)$, where $\sigma \geq \frac{\Delta q}{2\epsilon}(K + \sqrt{K^2 + 2\epsilon})$ and $K = Q^{-1}(\delta)$, satisfies (ϵ, δ)-DP.*

We define $\kappa_{\delta,\epsilon} = \frac{1}{2\epsilon}(K + \sqrt{K^2 + 2\epsilon})$, then the standard deviation σ in Theorem F.1 can be written as $\sigma(\delta, \epsilon) = \kappa_{\delta,\epsilon}\Delta q$. It can be shown that $\kappa_{\delta,\epsilon}$ behaves roughly as $O(\ln(1/\delta))^{1/2}/\epsilon$. For example, to ensure (ϵ, δ)-differential privacy with $\epsilon = \ln(2)$ and $\delta = 0.05$, the standard deviation of the injected Gaussian noise should be about 2.65 times the ℓ_1-sensitivity of q.

Theorem F.2. *The Gaussian Mechanism in Theorem F.1 is* $(\gamma, 2 \cdot Q(\frac{\gamma}{\sigma(\delta,\epsilon)}))$*-useful.*

Similar to our R^{2}DP Laplace mechanism, we can formulate an optimization problem for the R^{2}DP model using Gaussian mechanism. Therefore, using Theorems F.1 and F.2, we have the following.

Corollary F.3. *Denote by u, the set of parameters for a probability distribution* f_σ*. Then, the optimal usefulness of an R^{2}DP Gaussian mechanism utilizing* f_σ*, at each quadruplet* $(\epsilon, \delta, \Delta q, \gamma)$ *is*

$$\begin{aligned} U_f(\epsilon, \delta, \Delta q, \gamma) = \max_{u \in \mathbb{R}^{|u|}} \; & \left(1 - 2 \cdot \mathbb{E}_\sigma(Q(\tfrac{\gamma}{\sigma(\delta,\epsilon)}))\right) \\ & \textit{subject to} \\ \max_{\forall S \in \mathrm{R}} & \left\{ \frac{\mathbb{P}(\mathcal{M}_q(d,\sigma) \in S)}{\mathbb{P}(\mathcal{M}_q(d',\sigma) \in S)} \right\} = \epsilon, \\ & \mathbb{E}_\sigma(Q(\epsilon\sigma - \tfrac{1}{2\sigma})) = \delta \end{aligned} \tag{16}$$

F.2.2 R^{2}DP and Rényi Differential Privacy. Despite its notable advantages in numerous applications, the definition of (ϵ, δ)-differential privacy has the following two limitations.

First, (ϵ, δ)-differential privacy was applied to the analysis of the Gaussian mechanism [21]. In contrast to the Laplace mechanism (whose privacy guarantee is characterized tightly and accurately by ϵ-differential privacy), a single Gaussian mechanism satisfies a curve of $(\epsilon(\delta), \delta)$-differential privacy definitions [21]. Picking any one point on this curve may leave out important information about the mechanism's actual behavior [62].

Second, (ϵ, δ)-differential privacy also has limitations on the composition of differential privacy [61]. By relaxing the guarantee to (ϵ, δ)-differential privacy, advanced composition allows tighter analyses for compositions of (pure) differentially private mechanisms. Iterating this process, however, quickly leads to a combinatorial explosion of parameters, as each application of an advanced composition theorem leads to a wide selection of possibilities for $(\epsilon(\delta), \delta)$-differentially private guarantees.

To address these shortcomings, Rényi differential privacy was proposed as a natural relaxation of differential privacy in [62].

Definition F.2 ((α, ϵ)-RDP). *A randomized mechanism* $M : \mathrm{D} \times \Omega \rightarrow \mathrm{R}$ *is said to have* ϵ*-Rényi differential privacy of order* α*, or* (α, ϵ)*-RDP for short, if for if for all adjacent* $d, d' \in \mathrm{D}$*, we have* $D_\alpha(M(d)||M(d')) \leq \epsilon$*, where* $D_\alpha(\cdot)$ *is the (parameterized) Rényi divergence [73].*

Compared to (ϵ, δ)-differential privacy, Rényi differential privacy is a strictly stronger privacy definition. It offers an operationally convenient and quantitatively accurate way of tracking cumulative privacy loss throughout execution of a standalone differentially private mechanism and across many such mechanisms [62]. Next, we give the Rényi differential privacy guarantee of our R^{2}DP mechanism and show that the privacy loss of R^{2}DP under Rényi DP can significantly (asymptotically for small α) outperform Laplace, Gaussian and Random Response mechanisms.

Theorem F.4. *If real-valued query* q *has sensitivity* 1*, then the R^{2}DP mechanism* $\mathcal{M}_q$*, leveraging MGF* M*, satisfies*

$$\begin{cases} (\alpha, \frac{1}{\alpha-1}\log\left[\frac{\alpha M(\alpha-1)+(\alpha-1)M(-\alpha)}{2\alpha-1}\right]) - RDP. & if\ \alpha > 1 \\ (1, M'(0) + M(-1) - 1) - RDP. & if\ \alpha = 1 \end{cases}$$

Proof. The above RDP guarantee follows Corollary 2 in [62] on the RDP guarantee of the classic Laplace mechanism. In particular, the above equations are derived using the following substitutions $exp(t/b) \rightarrow M(t)$ and $1/b \rightarrow M'(0)$ due to the second-fold randomization of b. □

G OTHER APPLICATIONS OF R^{2}DP

R^{2}DP represents a very general concept which could potentially be applied in a broader range of contexts. In general, applying R^{2}DP to design more application-aware mechanisms may further improve the utility of many existing solutions, e.g., work [64] (Section 6 surveys other existing utility-maximizing schemes). We outline some of the possible applications as follows.

G.1 R^{2}DP and Composition

R^{2}DP may be applied for reducing the differential privacy leakage due to sequential or parallel querying over a dataset. In those scenarios, the objective will be to maximize the number of compositions under a specified ϵ-differential privacy constraint.

G.2 R^{2}DP and Local Differential Privacy

In this context, R^{2}DP can be regarded as a new randomized response model. In particular, the randomized response scheme presented in [76] can be produced using R^{2}DP for the Bernoulli distribution when $b_0 \rightarrow 0$ and $b_1 \rightarrow \infty$. Therefore, designing more efficient local differential privacy schemes using R^{2}DP is an interesting future direction.

G.3 R^{2}DP for Continual Observation Applications

Providing differential privacy guarantees on data streams represents another important future direction for R^{2}DP. As an example, the multi-input multi-output (MIMO) systems process streams of signals originated from many sensors capturing privacy-sensitive events about individuals, and statistics of interest need to be continuously published in real time [24, 53], e.g., privacy-preserving traffic monitoring over multi-lane roads [11]. In this context, R^{2}DP can leverage the constraint related to the number of inputs and the number of outputs (e.g., the sensitivity of the output of MIMO filter G with m inputs and p outputs is proportional to the $\mathcal{H}_2$ norm of G which itself is an increasing function of m and p [65]) into its model to build more efficient differentially private mechanisms for the MIMO scenarios.