Exactly Minimax-Optimal Locally Differentially Private Sampling

Hyun-Young Park
School of Electrical Engineering
KAIST
phy811@kaist.ac.kr

Shahab Asoodeh
Department of Computing and Software
McMaster University
asoodeh@mcmaster.ca

Si-Hyeon Lee
School of Electrical Engineering
KAIST
sihyeon@kaist.ac.kr

Abstract

The sampling problem under local differential privacy has recently been studied with potential applications to generative models, but a fundamental analysis of its privacy-utility trade-off (PUT) remains incomplete. In this work, we define the fundamental PUT of private sampling in the minimax sense, using the f-divergence between original and sampling distributions as the utility measure. We characterize the exact PUT for both finite and continuous data spaces under some mild conditions on the data distributions, and propose sampling mechanisms that are universally optimal for all f-divergences. Our numerical experiments demonstrate the superiority of our mechanisms over baselines, in terms of theoretical utilities for finite data space and of empirical utilities for continuous data space.

1 Introduction

Privacy leakage is a pressing concern in the realm of machine learning (ML), spurring extensive research into privacy protection techniques [1–4]. Among these, local differential privacy (LDP) [5] stands out as a standard model and has been deployed in industry, e.g., by Google [6, 7], Apple [8], Microsoft [9]. In the LDP framework, individual clients randomize their data on their own devices and send it to a potentially untrusted aggregator for analysis, thus preventing the user data from being inferred. However, this perturbation inherently diminishes data utility. Consequently, the central challenge in privacy mechanism design lies in optimizing utility while preserving the desired level of privacy protection. This goal involves characterizing the optimal balance between privacy parameter and utility, referred to as the privacy-utility trade-off (PUT). The analysis of the PUT and the proposal of privacy mechanisms have been actively conducted for various settings of statistical inference and machine learning [10–27].

Most research in this field focuses on scenarios where each client has only a single data point. However, there are increasingly more applications where each client has a large local dataset with multiple data records. One can formulate the privacy requirement in these cases by assuming that clients have datasets of the same size, generated independently from an underlying distribution [28–34]. This probabilistic assumption, however, restricts practical flexibility. The work [35] explored a scenario where clients have large datasets that may vary in size and seek to privately release another dataset that closely resembles their original dataset. In this scenario, local datasets can be represented by an empirical distribution, allowing each client to be seen as holding a probability distribution and

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

 https://doi.org/10.52202/079017-0329

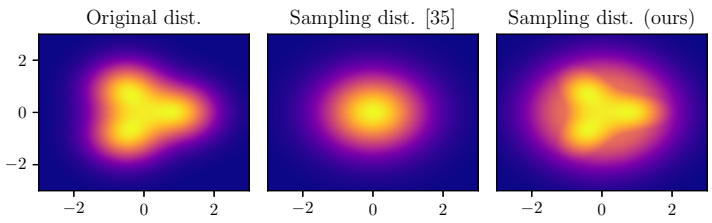

Figure 1: Original Gaussian ring distribution and the sampling distributions of the baseline [35] and our proposed mechanism for privacy budget $\epsilon = 0.5$. The implementation details are in Appendix F.

generating a private sample from it. This setup, which is called *private sampling*, is the main focus of this paper.

Private sampling has recently found applications in the private fine-tuning of large language models [36]. Additionally, private sampling is connected to the challenge of learning private generative models, a topic often explored in the central DP model [37–44]. While there exist studies on private generative models within the local model [45–48], all these works assume a single data point per client. Very recently, the work [49] considered a setup where each client holds a probability distribution, but in a different context of query estimation.

The private sampling mechanism in [35] can be described as follows. Initially, given a probability distribution P representing a local dataset, the mechanism assumes a *fixed* reference distribution. It then constructs what is termed the "relative mollifier", a closed ball centered around the reference distribution with a radius equivalent to half of the privacy budget, within the space of probability distributions. Subsequently, the mechanism computes the projection of P onto the relative mollifier, utilizing the Kullback-Leibler (KL) divergence. This projected distribution serves as the sampling distribution for generating a sample (see Section 2.3 for more details). However, this mechanism has a notable shortcoming: the sampling distribution is only locally optimal within the relative mollifier. This, in turn, implies that the optimality of the sampling distribution depends on the choice of the reference distribution. A more fundamentally intriguing goal would be to formulate and characterize the PUT without such an ambiguity in the choice of reference distribution.

In this paper, we establish the optimality of locally private sampling in the minimax sense, and identify optimal private samplers. Our primary contributions are summarized as follows:

- The fundamental PUT of private sampling is rigorously defined in terms of minimax utility, which is commonly used in the literature of private estimation [10, 11, 13, 30]. We impose some minimal assumptions on client's distributions as in [38] (which studies sampling under *central* DP, a weaker privacy model than the local model [50]). For utility measure, we use the *f-divergence* [51, 52] between the original and the sampling distributions, that includes KL divergence, total variation distance, squared Hellinger distance, and χ^2-divergence as special cases.
- We characterize the exact PUT for both finite and continuous data spaces, and present optimal sampling mechanisms achieving the PUT. Surprisingly, our mechanisms are *universally optimal* under any choice of f-divergence for utility measure.
- We numerically demonstrate that our proposed mechanism outperforms the baseline method presented in [35]. Specifically, for finite data spaces, we derive a closed-form expression for the utility of both our mechanism and the baseline, allowing for an exact comparison of their utilities. In the case of continuous data spaces, a closed-form expression for the baseline is not available, so we use empirical utility for the comparison. Figure 1 illustrates our proposed mechanism outputs a distribution closer to the original, than the baseline.

All codes for experiments and figures are attached as a supplementary material, and can be found at the online repository[1]. The instructions to reproduce the results in the paper are in Appendix H.

[1] `https://github.com/phy811/Optimal-LDP-Sampling`.

2 Problem Formulation

2.1 Notations and preliminaries

Notations. For a sample space $\mathcal{X}$, let $\mathcal{P}(\mathcal{X})$ be the set of all probability distributions on $\mathcal{X}$. For each $n \in \mathbb{N}$, let $\mathcal{C}(\mathbb{R}^n)$ be the set of all continuous probability distributions on $\mathbb{R}^n$. For each positive integer $k \in \mathbb{N}$, let $[k] := \{1, 2, \cdots, k\}$. For a subset $A \subset \mathcal{X}$, $\mathbb{1}_A : \mathcal{X} \to \{0, 1\}$ denotes the indicator function, defined as $\mathbb{1}_A(x) = 1$ for $x \in A$ and $\mathbb{1}_A(x) = 0$ for $x \notin A$. Also, for $s_1 \leq s_2$ and $x \in \mathbb{R}$, let $\mathrm{clip}(x; s_1, s_2) := \max\{s_1, \min\{s_2, x\}\}$. We refer to Appendix A for the rigorous measure-theoretic assumptions underlying the paper.

f-divergence. For a convex function $f : (0, \infty) \to \mathbb{R}$ satisfying $f(1) = 0$ and two probability distributions $P, Q \in \mathcal{P}(\mathcal{X})$ on the same sample space $\mathcal{X}$, let $D_f(P\|Q)$ denote the f-divergence [51, 52]. The general definition of f-divergence is given in Appendix B. For $\mathcal{X} = \mathbb{R}^n$ and $P, Q \in \mathcal{C}(\mathbb{R}^n)$ with $P \ll Q$ (that is, $P(A) = 0$ whenever $Q(A) = 0$), it is defined as

$$D_f(P\|Q) = \int_{x:q(x)>0} q(x) f\left(\frac{p(x)}{q(x)}\right) dx, \tag{1}$$

where p, q are pdfs of P, Q, respectively, and we define $f(0) = \lim_{x\to 0^+} f(x) \in (-\infty, \infty]$. For finite $\mathcal{X}$, we can replace the integral with the sum and replace p, q with P, Q. Several well-known distance measures between distributions are examples of f-divergence with different convex functions. For instance, KL divergence (relative entropy), total variation distance, squared Hellinger distance, and χ^2-divergence are f-divergences with $f(x) = x \log x$, $f(x) = |x - 1|/2$, $f(x) = (1 - \sqrt{x})^2$, and $f(x) = x^2 - 1$, respectively. Two important properties of general f-divergence are keys for this work. First, $D_f(P\|Q) \geq 0$, and equality holds if $P = Q$. Furthermore, we have

$$D_f(P\|Q) \leq M_f := \lim_{x\to 0+} f(x) + x f(1/x), \tag{2}$$

where equality holds if P and Q are mutually singular (that is, they have disjoint supports). For a more comprehensive list of such f-divergences and their properties, we refer the readers to [52].

We denote the KL divergence and the total variation distance as $D_{\mathrm{KL}}(P\|Q)$ and $D_{\mathrm{TV}}(P, Q)$, respectively. We note that the total variation distance is in fact a metric on $\mathcal{P}(\mathcal{X})$.

2.2 System model

Suppose a client has access to a distribution $P \in \mathcal{P}(\mathcal{X})$ over a sample space $\mathcal{X}$, and wants to produce a sample in $\mathcal{X}$ which looks like being drawn from P and to send it to a data curator. We assume that there are some constraints on the possible data distribution P, so that P is restricted to be in some subset $\tilde{\mathcal{P}} \subset \mathcal{P}(\mathcal{X})$, and both the client and the curator know $\mathcal{X}$ and $\tilde{\mathcal{P}}$. However, it is required that a sampled element does not leak the privacy about the original distribution P. For this purpose, the client and the curator agree a **private sampling mechanism $\mathbf{Q}$**, which is a conditional distribution from $\tilde{\mathcal{P}}$ to $\mathcal{X}$. After that, the client produces a sample following the distribution $\mathbf{Q}(\cdot|P)$. To guarantee the privacy protection, we impose $\mathbf{Q}$ to satisfy the local differential privacy (LDP) [10, 35].

Definition 2.1. Let $\epsilon > 0$. A private sampling mechanism $\mathbf{Q}$ is said to satisfy **ϵ-LDP**, or $\mathbf{Q}$ is an ϵ-LDP mechanism, if for any $P, P' \in \tilde{\mathcal{P}}$ and $A \subset \mathcal{X}$, we have

$$\mathbf{Q}(A|P) \leq e^{\epsilon} \mathbf{Q}(A|P'). \tag{3}$$

For convenience, for each $P \in \tilde{\mathcal{P}}$, let $\mathbf{Q}(P) \in \mathcal{P}(\mathcal{X})$ denote the distribution of X given P through $\mathbf{Q}$, that is $\mathbf{Q}(P)(A) = \mathbf{Q}(A|P)$ for each $A \subset \mathcal{X}$. In this way, we equivalently see $\mathbf{Q}$ as a function $\mathbf{Q} : \tilde{\mathcal{P}} \to \mathcal{P}(\mathcal{X})$. Let $\mathcal{Q}_{\mathcal{X},\tilde{\mathcal{P}},\epsilon}$ denote the set of all ϵ-LDP mechanisms $\mathbf{Q} : \tilde{\mathcal{P}} \to \mathcal{P}(\mathcal{X})$.

As the utility loss of the private sampling, we use the f-divergence between the original distribution and the sampling distribution, $D_f(P\|\mathbf{Q}(P))$. Since the sampling procedure can be performed across many clients who may have different data distributions, we measure the utility loss of $\mathbf{Q}$ by the **worst-case f-divergence**,

$$R_f(\mathbf{Q}) = \sup_{P\in\tilde{\mathcal{P}}} D_f(P\|\mathbf{Q}(P)). \tag{4}$$

https://doi.org/10.52202/079017-0329

Given $\mathcal{X}$, $\tilde{\mathcal{P}}$, ϵ, and f, our goal is to find the smallest possible worst-case f-divergence,

$$\mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f) = \inf_{\mathbf{Q} \in \mathcal{Q}_{\mathcal{X}, \tilde{\mathcal{P}}, \epsilon}} R_f(\mathbf{Q}), \tag{5}$$

and to find a mechanism $\mathbf{Q} \in \mathcal{Q}_{\mathcal{X}, \tilde{\mathcal{P}}, \epsilon}$ achieving it. We say that $\mathbf{Q} \in \mathcal{Q}_{\mathcal{X}, \tilde{\mathcal{P}}, \epsilon}$ is **optimal** for $(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon)$ under D_f if $R_f(\mathbf{Q}) = \mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f)$.

2.3 Related work

The most closely related work to our work is [35]. The system models are the same as this paper, except the formulation of PUT. They first *fix* a reference probability distribution $Q_0 \in \mathcal{P}(\mathcal{X})$, and only consider mechanisms $\mathbf{Q}$ satisfying $e^{-\epsilon/2} Q_0(A) \le \mathbf{Q}(A|P) \le e^{\epsilon/2} Q_0(A)$ for all $P \in \tilde{\mathcal{P}}$ and $A \subset \mathcal{X}$. In other words, let $\mathcal{M}_{\epsilon, Q_0} = \{Q \in \mathcal{P}(\mathcal{X}) : e^{-\epsilon/2} Q_0(A) \le Q(A) \le e^{\epsilon/2} Q_0(A), \quad \forall A \subset \mathcal{X}\}$. Then, they only consider $\mathbf{Q}$ such that $\mathbf{Q}(P) \in \mathcal{M}_{\epsilon, Q_0}$. Note that this guarantees ϵ-LDP. For each $P \in \tilde{\mathcal{P}}$, they sought to find $Q \in \mathcal{M}_{\epsilon, Q_0}$ which minimizes $D_{\mathrm{KL}}(P \| Q)$, and set this Q to be $\mathbf{Q}(P)$. First, they claimed to find a closed-form expression of such a minimizer Q for finite $\mathcal{X}$, given by

$$Q(x) = \mathrm{clip}\left(P(x)/r_P; e^{-\epsilon/2} Q_0(x), e^{\epsilon/2} Q_0(x)\right), \tag{6}$$

where $r_P > 0$ is a constant depending on P ensuring $\sum_{x \in \mathcal{X}} Q(x) = 1$. Second, they presented an algorithm, called Mollified Boosted Density Estimation (MBDE), to approximate the optimal solution for continuous $\mathcal{X}$. However, the utility varies over the choice of reference distribution Q_0, and they left the question of choosing a best Q_0 to achieve the best performance in both practice and theory. Moreover, we found that the closed-form in (6) is incomplete, because for some (P, Q_0), there may be no $r_P > 0$ such that the RHS of (6) does not sum to one. As an example, when P, Q_0 are point masses at different points, then we can easily see that the sum of (6) is $e^{-\epsilon/2}$ for any $r_P > 0$.

3 Main Results

3.1 Optimal private sampling over finite space

First, we consider the finite case, where $\mathcal{X} = [k]$ for some $k \in \mathbb{N}$. A natural setup for $\tilde{\mathcal{P}}$ is that $\tilde{\mathcal{P}} = \mathcal{P}([k])$, i.e. there is no restriction on the client distribution $P \in \mathcal{P}([k])$. In this case, we completely characterize the optimal worst-case f-divergence $\mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f)$ and find an optimal private sampling mechanism. Surprisingly, for each $k \in \mathbb{N}$ and $\epsilon > 0$, we found a single mechanism which is universally optimal for every f-divergence.

Theorem 3.1. *For each $k \in \mathbb{N}$, $\epsilon > 0$, and an f-divergence D_f, we have*

$$\mathcal{R}([k], \mathcal{P}([k]), \epsilon, f) = \frac{e^\epsilon}{e^\epsilon + k - 1} f\left(\frac{e^\epsilon + k - 1}{e^\epsilon}\right) + \frac{k-1}{e^\epsilon + k - 1} f(0). \tag{7}$$

Moreover, the mechanism $\mathbf{Q}^*_{k,\epsilon}$ constructed as below satisfies ϵ-LDP and is optimal for $(\mathcal{X} = [k], \tilde{\mathcal{P}} = \mathcal{P}([k]), \epsilon)$ under any D_f:

$$\mathbf{Q}^*_{k,\epsilon}(x|P) = \max\left(\frac{1}{r_P} P(x), \frac{1}{e^\epsilon + k - 1}\right) \quad \forall x \in [k], P \in \mathcal{P}([k]), \tag{8}$$

where $r_P > 0$ is a constant depending on P so that $\sum_{x=1}^k \mathbf{Q}^*_{k,\epsilon}(x|P) = 1$. Furthermore, r_P can be chosen such that $1 \le r_P \le (e^\epsilon + k - 1)/e^\epsilon$.

By definition, we have $\mathbf{Q}^*_{k,\epsilon}(x|P) \ge \frac{1}{e^\epsilon + k - 1}$ for all $x \in \mathcal{X}$. This also implies that $\mathbf{Q}^*_{k,\epsilon}(x|P) = 1 - \sum_{x' \in \mathcal{X} \setminus \{x\}} \mathbf{Q}^*_{k,\epsilon}(x'|P) \le 1 - \frac{k-1}{e^\epsilon + k - 1} = \frac{e^\epsilon}{e^\epsilon + k - 1}$. Hence, $\frac{1}{e^\epsilon + k - 1} \le \mathbf{Q}^*_{k,\epsilon}(x|P) \le \frac{e^\epsilon}{e^\epsilon + k - 1}$. This clearly implies that $\mathbf{Q}^*_{k,\epsilon}$ satisfies ϵ-LDP.

Behaviors of the optimal mechanism. Let us observe some behaviors of the proposed mechanism with respect to the system parameters, whose formal proofs are in Appendix E. We visualize how the mechanism $\mathbf{Q}^*_{k,\epsilon}$ works for different ϵ in Figure 2. Here, we write $\mathcal{R}$ to mean $\mathcal{R}([k], \mathcal{P}([k]), \epsilon, f)$

for simplicity. If $f(0) = \infty$, then $\mathcal{R} = \infty$, which means that $R_f(\mathbf{Q}) = \infty$ for any ϵ-LDP sampling mechanism $\mathbf{Q}$. (Such a phenomenon happens for general $(\mathcal{X}, \tilde{\mathcal{P}})$, whenever $\tilde{\mathcal{P}}$ contains two mutually singular distributions) Hence, from now on in this paragraph, we assume $f(0) < \infty$. We can observe that $\mathcal{R}$ is decreasing in ϵ and increasing in k. For a fixed k, we have $\mathcal{R} \to 0$ as $\epsilon \to \infty$, which makes sense since $\epsilon \to \infty$ corresponds to the non-private case. Also, as $\epsilon \to 0$, we have $\mathbf{Q}^*_{k,\epsilon}(x|P) \to 1/k$ for every $P \in \mathcal{P}([k])$ and $x \in [k]$, that is, $\mathbf{Q}^*_{k,\epsilon}(P)$ tends to the uniform distribution over $[k]$ for every $P \in \mathcal{P}([k])$. This fact can be also observed by Figure 2.

Remarks about the constant r_P. The value of r_P may not be unique, but the mechanism $\mathbf{Q}^*_{k,\epsilon}$ does not depend on the choice of r_P. To see this, let us fix P, and let $g_r(x) = \max\left(\frac{1}{r}P(x), \frac{1}{e^\epsilon+k-1}\right)$. Suppose that $\sum_{x=1}^k g_r(x) = \sum_{x=1}^k g_{r'}(x) = 1$ for $r \le r'$. Since $g_r(x) \ge g_{r'}(x)$ for each $x \in [k]$, the equality $\sum_{x=1}^k g_r(x) = \sum_{x=1}^k g_{r'}(x)$ implies that $g_r(x) = g_{r'}(x)$ for all $x \in [k]$. Hence $\mathbf{Q}^*_{k,\epsilon}$ is uniquely determined. Since $r \mapsto \sum_{x=1}^k g_r(x)$ is non-increasing and continuous, we can use the bisection method to find r_P. The 'Furthermore' part of the theorem statement precisely means that for $r = 1$ and $r = (e^\epsilon + k - 1)/e^\epsilon$, the value of $\sum_{x=1}^k g_r(x)$ is at least and at most 1, respectively, so that we can perform the bisection method with these two initial endpoints to find r such that $\sum_{x=1}^k g_r(x) = 1$.

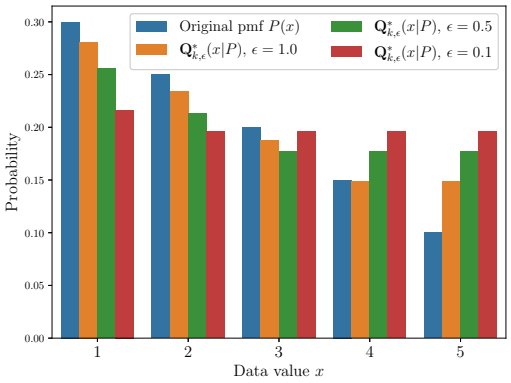

Figure 2: A visualization of the mechanism $\mathbf{Q}^*_{k,\epsilon}$

Comparison with the previous work [35]. The expression of the optimal mechanism in (8) is similar to (6), the expression of the KL divergence projection onto $\mathcal{M}_{\epsilon,Q_0}$ derived by Husain et al. [35]. Since $\frac{1}{e^\epsilon+k-1} \le \mathbf{Q}^*_{k,\epsilon}(x|P) \le \frac{e^\epsilon}{e^\epsilon+k-1}$, we can alternatively write $\mathbf{Q}^*_{k,\epsilon}(x|P) = \mathrm{clip}\left(\frac{1}{r_P}P(x); \frac{1}{e^\epsilon+k-1}, \frac{e^\epsilon}{e^\epsilon+k-1}\right)$. Hence, our optimal mechanism can be viewed as an instance of a generalized version of (6), where Q_0 is a positive measure, not necessarily a probability measure summing to one, given by $Q_0(x) = \frac{e^{\epsilon/2}}{e^\epsilon+k-1}$. A natural question is whether $\mathbf{Q}^*_{k,\epsilon}(P)$ is a projection of P onto $\mathcal{M}_{\epsilon,Q_0}$, that is $\mathbf{Q}^*_{k,\epsilon}(P)$ is a minimizer of $D_{\mathrm{KL}}(P\|Q)$ among $Q \in \mathcal{M}_{\epsilon,Q_0}$, where $\mathcal{M}_{\epsilon,Q_0}$ is similarly defined as in Section 2.3. As we shall discuss in Section 3.3, this statement is true —quite surprisingly— even when we replace D_{KL} with any other f-divergences. However, our analysis is more involved, as we need to show the optimality of the proposed mechanism over *any other possible mechanisms*, including minimizers with respect to other choices of Q_0. Also, in Section 5, we compare the worst-case f-divergence of our optimal mechanism with that of the mechanism proposed in [35] which restricts Q_0 to be a probability distribution.

3.2 Optimal private sampling over continuous space

Next, we consider the continuous case, where $\mathcal{X} = \mathbb{R}^n$ for some $n \in \mathbb{N}$. Some of the natural setups for $\tilde{\mathcal{P}}$ are (i) $\tilde{\mathcal{P}} = \mathcal{P}(\mathbb{R}^n)$, or (ii) $\tilde{\mathcal{P}} = \mathcal{C}(\mathbb{R}^n)$. We can also think some restrictive but still reasonable setups, such as the setups where (iii) $\tilde{\mathcal{P}}$ is the set of empirical distributions supported on some non-empty open subset of $\mathbb{R}^n$, or where (iv) $\tilde{\mathcal{P}}$ is the set of continuous distributions on $[-1, 1]^n$ having smooth pdf and zero mean. However, we show that for a general class of $\tilde{\mathcal{P}}$ including these four cases, any ϵ-LDP sampling mechanisms have the worst-case f-divergence equal to the maximum value M_f of the f-divergence defined in (2). In the following proposition, $\mathcal{X}$ may be a general sample space, not necessarily $\mathbb{R}^n$ or finite space. The proof is in Appendix D.

Proposition 3.2. *Suppose that $\tilde{\mathcal{P}}$ contains infinitely many distributions which are pairwise mutually singular. Then, for any $\epsilon > 0$, for any ϵ-LDP mechanism $\mathbf{Q} \in \mathcal{Q}_{\mathcal{X},\tilde{\mathcal{P}},\epsilon}$, and for any f-divergence, we have $R_f(\mathbf{Q}) = M_f$, where M_f is define at* (2).

https://doi.org/10.52202/079017-0329

Hence, we need to consider sufficiently regular but practical setups for $\tilde{\mathcal{P}}$.

Our setup. In this subsection, when $P, Q \in \mathcal{C}(\mathbb{R}^n)$, the corresponding small letters p, q denote their pdfs. In this paper, we consider the case that

$$\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,h} := \{P \in \mathcal{C}(\mathbb{R}^n) : c_1 h(x) \le p(x) \le c_2 h(x), \quad \forall x \in \mathbb{R}^n\} \tag{9}$$

for some pre-known $h : \mathcal{X} \to [0, \infty)$ such that $\int_{\mathbb{R}^n} h(x)dx < \infty$ and $c_2 > c_1 \ge 0$. Some of the sampling tasks and generative models in literature [53, 35] assume that the set of possible data distributions $\tilde{\mathcal{P}}$ satisfies $\tilde{\mathcal{P}} \subset \tilde{\mathcal{P}}_{c_1,c_2,h}$ for some c_1, c_2, h satisfying the aforementioned condition, hence (9) is a moderate assumption. One example is the sampling from a Gaussian mixture, which is one of the canonical sampling tasks in literature [53, 35]. Suppose that $\tilde{\mathcal{P}}$ consists of Gaussian mixtures, where each Gaussian has mean within a unit ball centered at the origin and has unit covariance. That is, $\tilde{\mathcal{P}} = \{\sum_{i=1}^k \lambda_i \mathcal{N}(\mu_i, I_n) : k \in \mathbb{N}, \lambda_i \ge 0, \sum_{i=1}^k \lambda_i = 1, \|\mu_i\| \le 1\}$, where I_n is the identity matrix of size $n \times n$. In this case, we can observe that $\tilde{\mathcal{P}} \subset \tilde{\mathcal{P}}_{0,1,h}$ for $h(x) = (2\pi)^{-n/2} \exp(-(\max(0, \|x\| - 1))^2/2)$, and it can be easily shown that $\int_{\mathbb{R}^n} h(x) < \infty$.

Without loss of generality, we may assume the following normalization condition on c_1, c_2, h, ϵ:

$$\int_{\mathbb{R}^n} h(x)dx = 1, \quad c_1 < 1 < c_2, \quad c_2 > c_1 e^\epsilon. \tag{10}$$

The reason is as follows. First, if any one of three inequalities $\int_{\mathbb{R}^n} h(x)dx > 0$, $c_1 \int_{\mathbb{R}^n} h(x)dx < 1$, and $c_2 \int_{\mathbb{R}^n} h(x)dx > 1$ is not satisfied, then $\tilde{\mathcal{P}}_{c_1,c_2,h}$ is either an empty set or a singleton that consists of a distribution having pdf $c_1 h(x)$ or $c_2 h(x)$, which makes the problem trivial. Hence we impose all of the three inequalities. Then, we can normalize c_1, c_2, h, to make $\int_{\mathbb{R}^n} h(x)dx = 1$ and $c_1 < 1 < c_2$. Furthermore, if $c_2 \le e^\epsilon c_1$, then for any $P_1, P_2 \in \tilde{\mathcal{P}}_{c_1,c_2,h}$, we have $p_1(x)/p_2(x) \le c_2/c_1 \le e^\epsilon$, hence we can easily observe that the mechanism $\mathbf{Q}$ defined as $\mathbf{Q}(P) = P$ for all $P \in \tilde{\mathcal{P}}_{c_1,c_2,h}$ satisfies ϵ-LDP and $R_f(\mathbf{Q}) = 0$, hence the problem also becomes trivial, giving $\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f) = 0$. Hence, we may assume (10).

Minimax utility and optimal mechanism. For the aforementioned setup, we can completely characterize $\mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f)$ and find a mechanism which is universally optimal for every f-divergence. The formula is similar to the discrete case, with a carefully chosen clipping bound.

Theorem 3.3. *For each $c_2 > c_1 \ge 0, \epsilon > 0$, and $h : \mathbb{R}^n \to [0, \infty)$ satisfying the normalization condition* (10)*, let us define the following constants determined by c_1, c_2, ϵ:*

$$b = \frac{c_2 - c_1}{(e^\epsilon - 1)(1 - c_1) + c_2 - c_1}, \quad r_1 = \frac{c_1}{b}, \quad r_2 = \frac{c_2}{be^\epsilon}. \tag{11}$$

Then, we have

$$\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f) = \frac{1 - r_1}{r_2 - r_1} f(r_2) + \frac{r_2 - 1}{r_2 - r_1} f(r_1). \tag{12}$$

Moreover, the mechanism $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$ constructed as below satisfies ϵ-LDP and is optimal for $(\mathcal{X} = \mathbb{R}^n, \tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon)$ under any D_f:

For each $P \in \tilde{\mathcal{P}}$, $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}(P) =: Q$ is defined as a continuous distribution with pdf

$$q(x) = \mathrm{clip}\left(\frac{1}{r_P} p(x); bh(x), be^\epsilon h(x)\right), \tag{13}$$

where $r_P > 0$ is a constant depending on P so that $\int_{\mathbb{R}^n} q(x)dx = 1$. Furthermore, r_P can be chosen such that $r_1 < r_P \le r_2$.

It is also clear that $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$ satisfies ϵ-LDP. Also, we note that $c_1 < b < 1 < be^\epsilon < c_2$ and $0 \le r_1 < 1 < r_2$, which is shown during the proof of Theorem 3.3 in Appendix C.

In practical scenario, it may be hard to expect $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,h}$ exactly, and we may only know $\tilde{\mathcal{P}} \subset \tilde{\mathcal{P}}_{c_1,c_2,h}$ for some c_1, c_2, h satisfying aforementioned conditions. In such case, we still propose to use $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$, and in Section 5, we numerically show that this proposed mechanism is better than previously proposed mechanism [35] in terms of the worst-case f-divergence.

Behavior of the optimal mechanism. We also observe some behaviors of the proposed mechanism with respect to the system parameters, whose formal proofs are in Appendix E. Again, we write $\mathcal{R}$ to mean $\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f)$ for simplicity. For a fixed (c_1, c_2), $\mathcal{R}$ is decreasing in ϵ. If $c_1 = 0$ (which implies $r_1 = 0$) and $f(0) = \infty$, then $\mathcal{R} = \infty$, which means $R_f(\mathbf{Q}) = \infty$ for any ϵ-LDP sampling mechanism $\mathbf{Q}$. For the behavior at $\epsilon \to \infty$ for a fixed (c_1, c_2), if $c_1 > 0$, then for sufficiently large ϵ, we have $c_1 e^\epsilon \geq c_2$, so we fall in the aforementioned trivial case that $\mathcal{R} = 0$. If $c_1 = 0$ and $f(0) < \infty$, then as $\epsilon \to \infty$, we have $\mathcal{R} \to 0$, which again corresponds to the non-private case.

Remarks on the constant r_P. By the same reason as in the finite space case, the value of r_P may not be unique, but $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$ does not depend on the choice of r_P, and r_P can be found by the bisection method with a numerical integration of (13). Note that the continuity of $r \mapsto \int g_r(x)dx, g_r(x) = \text{clip}\left(\frac{1}{r}p(x); bh(x), be^\epsilon h(x)\right)$, follows from the dominated convergence theorem [54] since we assume $\int_{\mathbb{R}^n} h(x)dx < \infty$. The meaning of 'Furthermore' part is also similar to the finite space case, that is, the value of $\int g_r(x)dx$ at $r = r_1$ and $r = r_2$ is at least and at most 1, respectively, so that we can perform the bisection method with initial endpoints (r_1, r_2) (When $r = 0$, we define $g_r(x) = be^\epsilon h(x)$ whenever $p(x) > 0$ and $g_r(x) = bh(x)$ whenever $p(x) = 0$). A corner case is that when $r_1 = 0$, the continuity of $r \mapsto \int g_r(x)dx$ does not suffice to guarantee the existence of strictly positive r such that $\int g_r(x)dx = 1$. However, in the proof, we actually show that $\int g_{r_1}(x)dx = 1$ implies $\int g_r(x)dx = 1$ for *every* $r \in (r_1, r_2]$, which especially implies that even when $r_1 = 0$, there is a *strictly positive* r such that $\int g_r(x)dx = 1$. This is the reason that we state the strict inequality $r_1 < r_P$.

3.3 Proof sketch of the theorems

The full proofs of the main theorems, Theorems 3.1 and 3.3, are presented in Appendix C. In Appendix C, we present a generalized theorem which includes Theorems 3.1 and 3.3 as special cases, where $\mathcal{X}$ can be a general sample space and $\tilde{\mathcal{P}}$ is similarly defined as in the continuous space case. The key idea for proofs and proposed mechanisms is to focus on the behavior of $\mathbf{Q}(P)$ when P is in an extreme case in $\tilde{\mathcal{P}}$. In finite space, point masses are extreme cases, and for continuous space with $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,h}$, the cases that $p(x) \in \{c_1 h(x), c_2 h(x)\}$ for all $x \in \mathcal{X}$ are the extreme cases. As implied by the proof, the worst-case f-divergence of the proposed optimal mechanism is attained when P is in the aforementioned extreme cases. Such an approach using extreme case is a frequently used technique in PUT analysis [55–58].

Our proof consists of two parts, the achievability part and the converse part. The achievability part is to show that the worst-case f-divergence $R_f(\mathbf{Q}^*)$ of our proposed mechanism $\mathbf{Q}^*$ is upper-bounded by the RHS of (7) or (12). The converse part is to show that $R_f(\mathbf{Q})$ of *any* ϵ-LDP mechanism $\mathbf{Q}$ is lower-bounded by the RHS of (7) or (12). From now, we briefly describe the proof idea of each part. Here, we omit the subscripts (k, ϵ) or (c_1, c_2, h, ϵ) for notational convenience.

Achievability part. Let $\mathcal{M} = \{\mathbf{Q}^*(P) : P \in \tilde{\mathcal{P}}\}$. For finite space, $\mathcal{M}$ consists of all distributions $Q \in \mathcal{P}([k])$ such that $\frac{1}{e^\epsilon + k - 1} \leq Q(x) \leq \frac{e^\epsilon}{e^\epsilon + k - 1}$ for every $x \in [k]$. For continuous space, $\mathcal{M}$ consists of all continuous distributions $Q \in \mathcal{C}(\mathbb{R}^n)$ whose pdf q satisfies $bh(x) \leq q(x) \leq be^\epsilon h(x)$ for every $x \in \mathbb{R}^n$.

We construct a mechanism $\mathbf{Q}^\dagger$ such that $R_f(\mathbf{Q}^\dagger)$ is upper-bounded by the RHS of (7) or (12). The construction is as follows. First, we set a reference distribution $\mu \in \mathcal{P}(\mathcal{X})$ and a constant $\gamma \in [0, 1]$ according to a certain rule specified in Appendix C. Then, for each given P, we generate a private sample by sampling from the original P with probability γ, and sampling from the reference distribution μ with probability $1 - \gamma$. In other words, we have $\mathbf{Q}^\dagger(P) = \gamma P + (1 - \gamma)\mu$. Our choice of μ and γ makes $\mathbf{Q}^\dagger(P) \in \mathcal{M}$ for every $P \in \tilde{\mathcal{P}}$, which especially implies that $\mathbf{Q}^\dagger$ also satisfies ϵ-LDP. Furthermore, we can find a bound on the ratio of the pmf or pdf for original distribution to that for sampling distribution. Then, invoking [59, Theorem 2.1], which bounds f-divergences given bounds on the ratio between pmf or pdfs, we show that $R_f(\mathbf{Q}^\dagger)$ is upper-bounded by the RHS of (7) or (12).

Next, we demonstrate a non-trivial generalization of the main result of [35] that $\mathbf{Q}^*(P)$ is the f-divergence projection of P onto $\mathcal{M}$ for $P \in \tilde{\mathcal{P}}$ and *for every f-divergence.*

 https://doi.org/10.52202/079017-0329

Proposition 3.4. *Assuming the setups of $(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon)$ in either Theorem 3.1 or 3.3, let $\mathbf{Q}^*$ denote the proposed mechanism $\mathbf{Q}^*_{k,\epsilon}$ or $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$. Also, let $\mathcal{M}$ be as described above. Then, for every $P \in \tilde{\mathcal{P}}$ and every f-divergence D_f, we have $D_f\left(P\|\mathbf{Q}^*(P)\right) = \inf_{Q\in\mathcal{M}} D_f\left(P\|Q\right)$.*

Notice that this proposition differs from [35] in that it holds for all general f-divergences (as opposed to only KL divergence) and general sample spaces, be it discrete or continuous. This result immediately yields $D_f\left(P\|\mathbf{Q}^*(P)\right) \le D_f\left(P\middle\|\mathbf{Q}^\dagger(P)\right)$, and hence $R_f(\mathbf{Q}^*) \le R_f(\mathbf{Q}^\dagger) \le$ (RHS of (7) or (12)). We remark that combining with the converse part (to be described below), it implies that the $\mathbf{Q}^\dagger$ is also optimal in our minimax sense. Nevertheless, $\mathbf{Q}^*$ outperforms $\mathbf{Q}^\dagger$ in that $D_f\left(P\|\mathbf{Q}^*(P)\right) \le D_f\left(P\middle\|\mathbf{Q}^\dagger(P)\right)$ for every $P \in \tilde{\mathcal{P}}$.

Remark 3.5. For the finite case, μ is the uniform distribution over $[k]$ and γ is taken in such a way that $\mathbf{Q}^\dagger$ satisfies ϵ-LDP tightly. In this case, an alternative way to implement $\mathbf{Q}^\dagger$ is as follows. For each given P, first we sample from P to get a raw sample and then apply the k-ary randomized response [60] to it.

Converse part. For each extreme P, let A_P be the "high probability set", defined as $A_P = \{x \in \mathcal{X} : P(x) = 1\}$ for finite space and $A_P = \{x \in \mathcal{X} : p(x) = c_2 h(x)\}$ for continuous space. Then, using the data processing inequality of f-divergence [61], we take a lower bound of $D_f\left(P\|\mathbf{Q}(P)\right)$ by the f-divergence between the distributions of $\mathbb{1}_{A_P}(X)$ for $X \sim P$ and that for $X \sim \mathbf{Q}(P)$. Such a lower bound becomes a decreasing function of $\mathbf{Q}(A_P|P)$ in a certain range. Then, we seek to find an upper bound on $\inf \mathbf{Q}(A_P|P)$ over extreme P, which gives a lower bound on $R_f(\mathbf{Q})$. This involves a novel combinatorial argument. We perform a "packing" of u copies of $\mathcal{X}$ by t subsets $A_1, \cdots, A_t$ with $A_i = A_{P_i}$ for some extreme P_i and appropriately chosen t and u. Then, we decompose the RHS of $u = \sum_{i=1}^{u} \mathbf{Q}(\mathcal{X}|P_i)$ by an appropriate partition of $\mathcal{X}$ involving A_i's and use the definition of ϵ-LDP to find an upper bound on $\inf \mathbf{Q}(A_P|P)$.

4 Discussions on the Proposed Mechanism

4.1 Effect of the r_P approximation error

In practice, it may not be possible to find the exact value of r_P such that the sum or integration of the RHS of (8) or (13) is 1. For the case of continuous space as in Theorem 3.3, one way to implement the proposed mechanism $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$ in practice is as follows. First, we fix parameters $\delta_1 \in [0, 1)$ and $\delta_2 \ge 0$ that quantify error tolerance. For a given P, we define $g_r(x) = \operatorname{clip}\left(\frac{1}{r}p(x); bh(x), be^\epsilon h(x)\right)$ and find $r_P > 0$ such that $\int_{\mathbb{R}^n} g_{r_P}(x)dx \in [1 - \delta_1, 1 + \delta_2]$; we delineate a numerical algorithm for this task in Section 3.2 based on the bisection method and a numerical integration method. Then, we get a private sample by sampling from the distribution with pdf $\hat{q}(x) = g_{r_P}(x)/\int_{\mathbb{R}^n} g_{r_P}(x)dx$. For the finite case as in Theorem 3.1, we can implement in the same way, except replacing the integral with the sum.

It is important to note that $\hat{q}(x) \in \left[\frac{bh(x)}{1+\delta_2}, \frac{be^\epsilon h(x)}{1-\delta_1}\right]$, indicating that the resulting $\mathbf{Q}^*_{k,\epsilon}$ and $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$ satisfy $\left(\epsilon + \log\frac{1+\delta_2}{1-\delta_1}\right)$-LDP as opposed to ϵ-LDP. Thus, the above implementation yields ϵ-LDP if it is used to implement $\mathbf{Q}^*_{k,\epsilon'}$ or $\mathbf{Q}^*_{c_1,c_2,h,\epsilon'}$, with $\epsilon' = \epsilon - \log\frac{1+\delta_2}{1-\delta_1}$ and sufficiently small δ_1, δ_2 such that $\epsilon' > 0$.

4.2 Continuity of the proposed mechanism

In some practical scenarios, the client may not have full access to their distribution P. One example is that the client can only access to samples from P. In such case, the client may first estimate the true distribution, and then perturb the estimated distribution through the optimal mechanism. The question is how the perturbation using the estimated distribution deviates from that using the true distribution. To answer this, we show that the proposed mechanism satisfies a pointwise Lipschitz property with respect to the total variation distance, and the Lipschitz constant is closely related to the factor r_P we introduce in Theorems 3.1 and 3.3.

Proposition 4.1. *Assuming the setups of $(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon)$ in either Theorem 3.1 or 3.3, let $\mathbf{Q}^*$ denote the proposed mechanism $\mathbf{Q}^*_{k,\epsilon}$ or $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$. Then, for any $P, P' \in \tilde{\mathcal{P}}$, we have*

$$D_{\mathrm{TV}}\left(\mathbf{Q}^*(P), \mathbf{Q}^*(P')\right) \leq \frac{2}{\max(r_P, r_{P'})} D_{\mathrm{TV}}\left(P, P'\right) \tag{14}$$

where $r_P > 0$ is as in Theorem 3.1 or 3.3.

This guarantees that for each given true P and given $\delta > 0$, whenever the approximated P' satisfies $D_{\mathrm{TV}}\left(P, P'\right) \leq \delta r_P/2$, the perturbed distribution $\mathbf{Q}^*(P')$ satisfies $D_{\mathrm{TV}}\left(\mathbf{Q}^*(P), \mathbf{Q}^*(P')\right) \leq \delta$. In theoretical perspective, this proposition implies that $\mathbf{Q}^*$ is continuous when $\tilde{\mathcal{P}}$ and $\mathcal{P}(\mathcal{X})$ are endowed with the metric topology from the total variation distance. The proof of Proposition 4.1 is in Appendix D.

5 Numerical Results

In this section, we numerically compare the worst-case f-divergence of our proposed mechanism with that of the previously proposed sampling mechanism. To the best of our knowledge, the only work about the private sampling under LDP is [35], hence we set the baseline as the mechanism proposed in [35]. In all the cases, we perform the comparison across three canonical f-divergences: KL divergence, total variation distance, and squared Hellinger distance, as well as across five values of ϵ: 0.1, 0.5, 1, 2, and 5.

5.1 Comparison for finite data space

In this subsection, we compare the mechanisms in the finite space, $\mathcal{X} = [k]$ and $\tilde{\mathcal{P}} = \mathcal{P}([k])$. As mentioned in Sections 2.3 and 3.1, the baseline mechanism has a hyper-parameter, a reference probability distribution $Q_0 \in \mathcal{P}(\mathcal{X})$. We set the baseline as a generalized f-divergence projection onto the relative mollifier. That is, for each given f-divergence, we set the baseline to satisfy $\mathbf{Q}(P) \in \arg\min_{Q \in \mathcal{M}_{\epsilon,Q_0}} D_f\left(P \| Q\right)$, where $\mathcal{M}_{\epsilon,Q_0}$ is defined in Section 2.3. As expected by symmetry, for any f, choosing Q_0 to be the uniform distribution minimizes the worst-case f-divergence $R_f(\mathbf{Q})$ among all choices of Q_0 for the baseline. Also, even though we do not obtain the closed-form expression of $\mathbf{Q}(x|P)$ for the baseline, we obtain the value of $R_f(\mathbf{Q})$ when Q_0 is the uniform distribution. The proof of this fact, together with the precise value of $R_f(\mathbf{Q})$ for uniform Q_0, is in Appendix F.1. Hence, we always set Q_0 to be the uniform distribution in the result about the baseline. Since we have the precise values of $R_f(\mathbf{Q})$ for both our proposed mechanism and the baseline, we plot such values of $R_f(\mathbf{Q})$ in Figure 3. For simplicity, we only provide the plot for $k = 10$. More plots for some other k's can be found in Appendix G. As shown by the figure, the proposed mechanism has lower worst-case f-divergence than the baseline for all choices of f-divergences and ϵ in the experiment, with significant gap in medium privacy regime $\epsilon \in [0.5, 2]$.

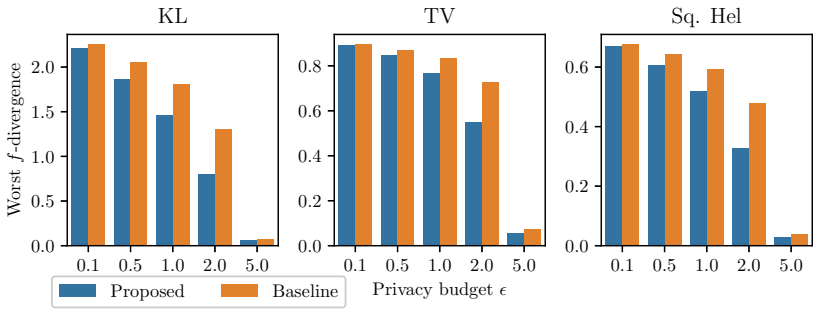

Figure 3: Theoretical worst-case f-divergences of proposed and previously proposed baseline mechanisms (with uniform Q_0) over finite space ($k = 10$)
(Left: KL divergence, Center: Total variation distance, Right: Squared Hellinger distance)

https://doi.org/10.52202/079017-0329

5.2 Comparison for 1D Gaussian mixture

In this subsection, we conduct an experiment to compare the mechanisms when the client distributions are Gaussian mixtures over a real line $\mathcal{X} = \mathbb{R}$, which is an instance of a continuous space case. We consider the case that each client has a Gaussian mixture distribution in $\mathbb{R}$, where each Gaussian has a mean bounded by 1 and has a unit variance. To avoid arbitrarily large number of Gaussian distributions to be mixed, we set an upper bound K of the number of Gaussian distributions to be mixed per client. Also, to make the numerical integration tractable, we truncate the domain of the distributions to lie inside an interval $[-4, 4]$. Unlike the finite space case, there is no known closed-form expression of the worst-case f-divergence for the mechanism in [35]. The set of Gaussian mixtures is not exactly of the form $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,h}$, hence our proposed mechanism also does not have a known closed-form expression of the worst-case f-divergence. Hence, instead, we compare the mechanisms by an empirical worst-case f-divergence.

For an experiment, we randomly construct N Gaussian mixture distributions $P_1, P_2, \cdots, P_N \in \tilde{\mathcal{P}}$, where each P_j is generated independently according to some rules specified in Appendix F.2. After that, we plot the value of the empirical maximum f-divergence $\max_{j\in[N]} D_f\left(P_j \| \mathbf{Q}(P_j)\right)$ for the baseline and our proposed $\mathbf{Q}$. For the baseline mechanism, we use MBDE with the same hyperparameter setup as [35, Section 5], except a slight modification of the reference distribution to consider the truncation of the domain. The implementation details are provided in Appendix F.2.

In Figure 4, we present the result for $N = 100$ and $K = 10$. We can see that the proposed mechanism has much lower worst-case f-divergence than the baseline for all choices of f-divergences and ϵ.

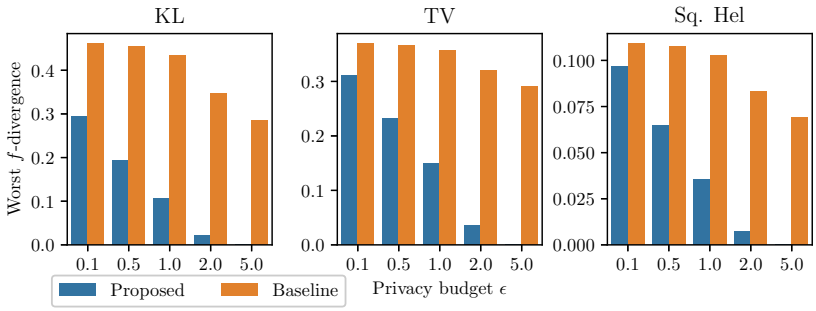

Figure 4: Empirical worst-case f-divergences of proposed and baseline mechanisms over 100 experiments of 1D Gaussian mixture
(Left: KL divergence, Center: Total variation distance, Right: Squared Hellinger distance)

6 Conclusion

In this paper, we characterized the optimal privacy-utility trade-off for the private sampling under LDP and found the optimal private sampling mechanism in terms of the minimax f-divergence between original and sampling distributions, for both finite and continuous data spaces. Compared to the previous work [35] based on relative mollifier with arbitrarily chosen reference distribution, our work characterizes PUT without dependency on external information other than the original distribution, and it is shown that the mechanism we found is universally optimal under any f-divergence.

For future works, there may be other $\tilde{\mathcal{P}}$ and other measures of utility more appropriate to practical scenarios, which are not handled in this paper. For example, f-divergence may be an inappropriate utility loss because it only depends on σ-algebra structure and does not consider additional information about geometry of $\mathcal{X}$, such as underlying metric on $\mathcal{X}$. Using utility measures involving the geometry, such as Wasserstein distance [62, 63, 49], may be more appropriate for some scenarios. Also, we can consider the Bayesian approach instead of the worst-case approach. Furthermore, we only consider the task of releasing a single sample per client in this paper. We may also consider the case of releasing multiple samples per client, rather than a single sample.

The limitations and broader impacts of this work are in Appendices I and J, respectively.

Acknowledgments and Disclosure of Funding

We thank the anonymous reviewers for their valuable discussions and comments, particularly regarding the identification of an alternative optimal mechanism, the advantage of the proposed mechanism over the alternative, and the effect of the approximation error on r_P.

This work was supported in part by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIT) under Grant 2022R1A2C2092151 and in part by Institute of Information & Communications Technology Planning & Evaluation (IITP) under 6G·Cloud Research and Education Open Hub (IITP-2024-RS-2024-00428780) grant funded by the Korea government (MSIT).

References

[1] Milad Nasr, Nicholas Carlini, Jonathan Hayase, Matthew Jagielski, A. Feder Cooper, Daphne Ippolito, Christopher A. Choquette-Choo, Eric Wallace, Florian Tramèr, and Katherine Lee. Scalable Extraction of Training Data from (Production) Language Models. *arXiv preprint arXiv:2311.17035*, November 2023.

[2] Xiaodong Wu, Ran Duan, and Jianbing Ni. Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence*, 2(2):102–115, March 2024. ISSN 29497159. doi: 10.1016/j.jiixd.2023.10.007.

[3] Biwei Yan, Kun Li, Minghui Xu, Yueyan Dong, Yue Zhang, Zhaochun Ren, and Xiuzhen Cheng. On Protecting the Data Privacy of Large Language Models (LLMs): A Survey. *arXiv preprint arXiv:2403.05156*, March 2024.

[4] Zhangheng Li, Junyuan Hong, Bo Li, and Zhangyang Wang. Shake to Leak: Fine-tuning Diffusion Models Can Amplify the Generative Privacy Risk. In *2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML)*, pages 18–32, April 2024. doi: 10.1109/SaTML59370.2024.00010.

[5] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What Can We Learn Privately? *SIAM Journal on Computing*, 40(3):793–826, January 2011. ISSN 0097-5397. doi: 10.1137/090756090.

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

[7] Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, page 441–459. Association for Computing Machinery, 2017. ISBN 9781450350853.

[8] Differential privacy team Apple. Learning with privacy at scale. *Technical report, Apple*, 2017.

[9] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting Telemetry Data Privately. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[10] John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Local Privacy and Statistical Minimax Rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438, October 2013. doi: 10.1109/FOCS.2013.53.

[11] Min Ye and Alexander Barg. Optimal Schemes for Discrete Distribution Estimation Under Locally Differential Privacy. *IEEE Transactions on Information Theory*, 64(8):5662–5676, August 2018. ISSN 0018-9448, 1557-9654. doi: 10.1109/TIT.2018.2809790.

[12] Abhishek Bhowmick, John Duchi, Julien Freudiger, Gaurav Kapoor, and Ryan Rogers. Protection Against Reconstruction and Its Applications in Private Federated Learning. *arXiv preprint arXiv:1812.00984*, June 2019.

[13] Hilal Asi, Vitaly Feldman, and Kunal Talwar. Optimal Algorithms for Mean Estimation under Local Differential Privacy. In *Proceedings of the 39th International Conference on Machine Learning*, pages 1046–1056. PMLR, June 2022.

 https://doi.org/10.52202/079017-0329

[14] Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, and Adam Sealfon. On computing pairwise statistics with local differential privacy. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 27129–27146. Curran Associates, Inc., 2023.

[15] Bonwoo Lee, Jeongyoun Ahn, and Cheolwoo Park. Minimax risks and optimal procedures for estimation under functional local differential privacy. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 57964–57975. Curran Associates, Inc., 2023.

[16] Hilal Asi, Vitaly Feldman, Jelani Nelson, Huy Nguyen, and Kunal Talwar. Fast optimal locally private mean estimation via random projections. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 16271–16282. Curran Associates, Inc., 2023.

[17] Berivan Isik, Wei-Ning Chen, Ayfer Ozgur, Tsachy Weissman, and Albert No. Exact optimality of communication-privacy-utility tradeoffs in distributed mean estimation. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 37761–37785. Curran Associates, Inc., 2023.

[18] Raef Bassily, Adam Smith, and Abhradeep Thakurta. Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 464–473, October 2014. doi: 10.1109/FOCS.2014.56.

[19] Stacey Truex, Ling Liu, Ka-Ho Chow, Mehmet Emre Gursoy, and Wenqi Wei. LDP-Fed: Federated learning with local differential privacy. In *Proceedings of the Third ACM International Workshop on Edge Systems, Analytics and Networking*, EdgeSys '20, pages 61–66, New York, NY, USA, May 2020. Association for Computing Machinery. ISBN 978-1-4503-7132-2. doi: 10.1145/3378679.3394533.

[20] Badih Ghazi, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Avinash V. Varadarajan, and Chiyuan Zhang. Regression with Label Differential Privacy. *arXiv preprint arXiv:2212.06074*, October 2023.

[21] Vitaly Feldman, Ilya Mironov, Kunal Talwar, and Abhradeep Thakurta. Privacy Amplification by Iteration. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 521–532, October 2018. doi: 10.1109/FOCS.2018.00056.

[22] Mengchu Li, Tom Berrett, and Yi Yu. Network change point localisation under local differential privacy. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 15013–15026. Curran Associates, Inc., 2022.

[23] Shuzhen Chen, Dongxiao Yu, Yifei Zou, Jiguo Yu, and Xiuzhen Cheng. Decentralized Wireless Federated Learning With Differential Privacy. *IEEE Transactions on Industrial Informatics*, 18(9):6273–6282, September 2022. ISSN 1941-0050. doi: 10.1109/TII.2022.3145010.

[24] Youssef Allouah, Rachid Guerraoui, Nirupam Gupta, Rafael Pinot, and John Stephan. On the Privacy-Robustness-Utility Trilemma in Distributed Learning. In *Proceedings of the 40th International Conference on Machine Learning*, pages 569–626. PMLR, July 2023.

[25] Chuan Guo, Kamalika Chaudhuri, Pierre Stock, and Michael Rabbat. Privacy-Aware Compression for Federated Learning Through Numerical Mechanism Design. In *Proceedings of the 40th International Conference on Machine Learning*, pages 11888–11904. PMLR, July 2023.

[26] Yi Liu, Qirui Hu, Lei Ding, and Linglong Kong. Online Local Differential Private Quantile Inference via Self-normalization. In *Proceedings of the 40th International Conference on Machine Learning*, pages 21698–21714. PMLR, July 2023.

[27] Jin Sima, Changlong Wu, Olgica Milenkovic, and Wojciech Szpankowski. Online Distribution Learning with Local Privacy Constraints. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, pages 460–468. PMLR, April 2024.

[28] Antonious M. Girgis, Deepesh Data, and Suhas Diggavi. Distributed User-Level Private Mean Estimation. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2196–2201, June 2022. doi: 10.1109/ISIT50566.2022.9834713.

[29] Ruiquan Huang, Huanyu Zhang, Luca Melis, Milan Shen, Meisam Hejazinia, and Jing Yang. Federated Linear Contextual Bandits with User-level Differential Privacy. In *Proceedings of the 40th International Conference on Machine Learning*, pages 14060–14095. PMLR, July 2023.

[30] Jayadev Acharya, Yuhan Liu, and Ziteng Sun. Discrete Distribution Estimation under User-level Local Differential Privacy. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, pages 8561–8585. PMLR, April 2023.

[31] Raef Bassily and Ziteng Sun. User-level Private Stochastic Convex Optimization with Optimal Rates. In *Proceedings of the 40th International Conference on Machine Learning*, pages 1838–1851. PMLR, July 2023.

[32] Yulian Mao, Qingqing Ye, Haibo Hu, Qi Wang, and Kai Huang. PrivShape: Extracting Shapes in Time Series under User-Level Local Differential Privacy. *arXiv preprint arXiv:2404.03873*, April 2024.

[33] Kangkang Sun, Jun Wu, Ali Kashif Bashir, Jianhua Li, Hansong Xu, Qianqian Pan, and Yasser D. Al-Otaibi. Personalized Privacy-Preserving Distributed Artificial Intelligence for Digital-Twin-Driven Vehicle Road Cooperation. *IEEE Internet of Things Journal*, pages 1–1, 2024. ISSN 2327-4662. doi: 10.1109/JIOT.2024.3389656.

[34] Alexander Kent, Thomas B. Berrett, and Yi Yu. Rate Optimality and Phase Transition for User-Level Local Differential Privacy. *arXiv preprint arXiv:2405.11923*, May 2024.

[35] Hisham Husain, Borja Balle, Zac Cranko, and Richard Nock. Local Differential Privacy for Sampling. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 3404–3413. PMLR, June 2020.

[36] James Flemings, Meisam Razaviyayn, and Murali Annavaram. Differentially Private Next-Token Prediction of Large Language Models. *arXiv preprint arXiv:2403.15638*, April 2024.

[37] Sofya Raskhodnikova, Satchit Sivakumar, Adam Smith, and Marika Swanberg. Differentially private sampling from distributions. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 28983–28994. Curran Associates, Inc., 2021.

[38] Badih Ghazi, Xiao Hu, Ravi Kumar, and Pasin Manurangsi. On differentially private sampling from gaussian and product distributions. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 77783–77809. Curran Associates, Inc., 2023.

[39] Hamid Ebadi, Thibaud Antignac, and David Sands. Sampling and partitioning for differential privacy. In *2016 14th Annual Conference on Privacy, Security and Trust (PST)*, pages 664–673, February 2016. doi: 10.1109/PST.2016.7906954.

[40] Nazmiye Ceren Abay, Yan Zhou, Murat Kantarcioglu, Bhavani Thuraisingham, and Latanya Sweeney. Privacy Preserving Synthetic Data Release Using Deep Learning. In Michele Berlingerio, Francesco Bonchi, Thomas Gärtner, Neil Hurley, and Georgiana Ifrim, editors, *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science, pages 510–526, Cham, 2019. Springer International Publishing. ISBN 978-3-030-10925-7. doi: 10.1007/978-3-030-10925-7_31.

[41] Liyang Xie, Kaixiang Lin, Shu Wang, Fei Wang, and Jiayu Zhou. Differentially Private Generative Adversarial Network. *arXiv preprint arXiv:1802.06739*, February 2018.

[42] Bangzhou Xin, Wei Yang, Yangyang Geng, Sheng Chen, Shaowei Wang, and Liusheng Huang. Private FL-GAN: Differential Privacy Synthetic Data Generation Based on Federated Learning. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2927–2931, May 2020. doi: 10.1109/ICASSP40776.2020.9054559.

[43] Reihaneh Torkzadehmahani, Peter Kairouz, and Benedict Paten. DP-CGAN: Differentially Private Synthetic Data and Label Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[44] Yi Liu, Jialiang Peng, James J.Q. Yu, and Yi Wu. PPGAN: Privacy-Preserving Generative Adversarial Network. In *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 985–989, February 2019. doi: 10.1109/ICPADS47876.2019.00150.

[45] Teddy Cunningham, Konstantin Klemmer, Hongkai Wen, and Hakan Ferhatosmanoglu. GeoPointGAN: Synthetic Spatial Data with Local Label Differential Privacy. *arXiv preprint arXiv:2205.08886*, May 2022.

[46] Hua Zhang, Kaixuan Li, Teng Huang, Xin Zhang, Wenmin Li, Zhengping Jin, Fei Gao, and Minghui Gao. Publishing locally private high-dimensional synthetic data efficiently. *Information Sciences*, 633:343–356, July 2023. ISSN 0020-0255. doi: 10.1016/j.ins.2023.03.014.

 https://doi.org/10.52202/079017-0329

[47] Hisaichi Shibata, Shouhei Hanaoka, Yang Cao, Masatoshi Yoshikawa, Tomomi Takenaga, Yukihiro Nomura, Naoto Hayashi, and Osamu Abe. Local Differential Privacy Image Generation Using Flow-Based Deep Generative Models. *Applied Sciences*, 13(18):10132, January 2023. ISSN 2076-3417. doi: 10.3390/app131810132.

[48] Hansle Gwon, Imjin Ahn, Yunha Kim, Hee Jun Kang, Hyeram Seo, Heejung Choi, Ha Na Cho, Minkyoung Kim, JiYe Han, Gaeun Kee, Seohyun Park, Kye Hwa Lee, Tae Joon Jun, and Young-Hak Kim. LDP-GAN : Generative adversarial networks with local differential privacy for patient medical records synthesis. *Computers in Biology and Medicine*, 168:107738, January 2024. ISSN 0010-4825. doi: 10.1016/j.compbiomed.2023.107738.

[49] Jacob Imola, Amrita Roy Chowdhury, and Kamalika Chaudhuri. Metric Differential Privacy at the User-Level. *arXiv preprint arXiv:2405.02665*, May 2024.

[50] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating Noise to Sensitivity in Private Data Analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, Lecture Notes in Computer Science, pages 265–284, Berlin, Heidelberg, 2006. Springer. ISBN 978-3-540-32732-5. doi: 10.1007/11681878_14.

[51] Igal Sason. On f-Divergences: Integral Representations, Local Behavior, and Inequalities. *Entropy*, 20(5): 383, May 2018. ISSN 1099-4300. doi: 10.3390/e20050383.

[52] Igal Sason and Sergio Verdú. f-Divergence Inequalities. *IEEE Transactions on Information Theory*, 62 (11):5973–6006, January 2016. ISSN 1557-9654. doi: 10.1109/TIT.2016.2603151.

[53] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144, October 2020. ISSN 0001-0782. doi: 10.1145/3422622.

[54] Elias M. Stein and Rami Shakarchi. *Real Analysis: Measure Theory, Integration, and Hilbert Spaces*. Princeton University Press, 2005. ISBN 978-0-691-11386-9. doi: 10.2307/j.ctvd58v18.

[55] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal Mechanisms for Local Differential Privacy. *Journal of Machine Learning Research*, 17(17):1–51, 2016. ISSN 1533-7928.

[56] Naoise Holohan, Douglas J. Leith, and Oliver Mason. Extreme points of the local differential privacy polytope. *Linear Algebra and its Applications*, 534:78–96, December 2017. ISSN 0024-3795. doi: 10.1016/j.laa.2017.08.011.

[57] Ankit Pensia, Amir R. Asadi, Varun Jog, and Po-Ling Loh. Simple Binary Hypothesis Testing under Local Differential Privacy and Communication Constraints. *arXiv preprint arXiv:2301.03566*, December 2023.

[58] Seung-Hyun Nam, Vincent Y. F. Tan, and Si-Hyeon Lee. Optimal Private Discrete Distribution Estimation With 1-bit Communication. *IEEE Transactions on Information Forensics and Security*, 19:6514–6528, 2024. ISSN 1556-6021. doi: 10.1109/TIFS.2024.3419721.

[59] Andrew Rukhin. Information-type divergence when the likelihood ratios are bounded. *Applicationes Mathematicae*, 4(24):415–423, 1997. ISSN 1233-7234.

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

[61] F. Liese and I. Vajda. On Divergences and Informations in Statistics and Information Theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, October 2006. ISSN 1557-9654. doi: 10.1109/TIT.2006.881731.

[62] Cédric Villani. *Optimal Transport: Old and New*. Number 338 in Grundlehren Der Mathematischen Wissenschaften. Springer, Berlin, 2009. ISBN 978-3-540-71049-3.

[63] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. *arXiv preprint arXiv:1701.07875*, December 2017.

[64] Stephen P. Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK ; New York, 2004. ISBN 978-0-521-83378-3.

[65] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970. ISBN 978-0-691-01586-6.

[66] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21 (6):1087–1092, June 1953. ISSN 0021-9606. doi: 10.1063/1.1699114.

[67] Christian P. Robert. The Metropolis-Hastings algorithm. *arXiv preprint arXiv:1504.01896*, January 2016.

[68] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The Composition Theorem for Differential Privacy. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1376–1385. PMLR, June 2015.

[69] Ryan M Rogers, Aaron Roth, Jonathan Ullman, and Salil Vadhan. Privacy Odometers and Filters: Pay-as-you-Go Composition. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016.

 https://doi.org/10.52202/079017-0329

A Assumptions about the Measure Theory

The appendices assume the familiarity with the basic measure theory and real analysis. We refer the standard textbooks about the measure theory and real analysis, e.g. [54].

Throughout the main paper and the appendices, we assume the followings:

- For each sample space $\mathcal{X}$, a σ-algebra on $\mathcal{X}$ is implicitly given. Unless mentioned otherwise,
  - the discrete σ-algebra is given for finite $\mathcal{X}$, and
  - the Borel σ-algebra is given for $\mathcal{X} = \mathbb{R}^n$.
- A "subset" of $\mathcal{X}$ always means a "measurable subset", and similarly $A \subset \mathcal{X}$ always means A is a measurable subset.
- The "continuous" distribution precisely means the "absolutely continuous" distribution (with respect to the Lebesgue measure).

In the appendices, we also introduce the following notations:

- For finite $\mathcal{X}$, $\#$ denotes the counting measure.
- For $\mathcal{X} = \mathbb{R}^n$, m denotes the Lebesgue measure.

B General Definition and More Properties of f-divergences

In this appendix, we review the general definition and additional properties about the f-divergences which are important in our analysis. Let a convex function $f : (0, \infty) \to \mathbb{R}$ with $f(1) = 0$ be given. For $P, Q \in \mathcal{P}(\mathcal{X})$, not necessarily $P \ll Q$, we first take a dominating measure μ on $\mathcal{X}$ such that $P, Q \ll \mu$ (e.g., $\mu = P + Q$), and let $p = dP/d\mu$, $q = dQ/d\mu$. The f-divergence $D_f(P\|Q)$ is defined as

$$D_f(P\|Q) = \int q(x) f\left(\frac{p(x)}{q(x)}\right) d\mu(x). \tag{15}$$

We note that (15) is invariant under the choice of the dominating measure μ.

Since it is possible that $p(x) = 0$ or $q(x) = 0$, we need to define what is the value of the expression $qf(p/q)$ for $p = 0$ or $q = 0$ [51, 52]. It is well-known that any convex function $f : (0, 1) \to \mathbb{R}$ is continuous, and the function

$$f^\star(x) = xf(1/x) \tag{16}$$

is also convex on $(0, 1)$. Furthermore, any convex function $f : (0, 1) \to \mathbb{R}$ with $f(1) = 0$ has a limit $\lim_{x\to 0+} f(x)$ in $\mathbb{R} \cup \{+\infty\}$. Hence, we have the continuous extension $f : [0, \infty) \to \mathbb{R} \cup \{+\infty\}$ by setting $f(0) = \lim_{x\to 0+} f(x)$, which is proper convex and continuous. By the same way, we have the continuous extension $f^\star : [0, \infty) \to \mathbb{R} \cup \{+\infty\}$. Using these extensions, we define $0f(0/0) = 0$, $qf(0/q) = qf(0)$ for $q > 0$, and $0f(p/0) = pf^\star(0)$ for $p > 0$. Especially, if $f^\star(0) = \infty$, then $D_f(P\|Q) = \infty$ whenever $P \ll Q$ does not hold. (This is the case for KL divergence and χ^2 divergence for examples) Similarly, if $f(0) = \infty$, then $D_f(P\|Q) = \infty$ whenever $Q \ll P$ does not hold. Also, the maximum value of the f-divergence M_f presented in (2) can be written as $M_f = f(0) + f^\star(0)$.

The following additional properties of f-divergences are important in our analysis. [61]

Theorem B.1. *Any f-divergence is jointly convex, that is for any $P_1, P_2, Q_1, Q_2 \in \mathcal{P}(\mathcal{X})$ and $0 \le \lambda \le 1$, we have $D_f(\lambda P_1 + (1-\lambda)P_2 \| \lambda Q_1 + (1-\lambda)Q_2) \le \lambda D_f(P_1\|Q_1) + (1-\lambda) D_f(P_2\|Q_2)$.*

Theorem B.2 (Data-Processing Inequality)**.** *Let M be a conditional distribution (Markov kernel) from $\mathcal{X}$ to $\mathcal{Y}$. For given $P_1, P_2 \in \mathcal{P}(\mathcal{X})$, let $Q_1, Q_2 \in \mathcal{P}(\mathcal{Y})$ be the push-forward measure of P_1, P_2 through M, respectively. Then for any f-divergence, we have $D_f(P_1\|P_2) \ge D_f(Q_1\|Q_2)$.*

Also, we present several equivalent expressions for the total variation distance [51, 52], which are used in Appendix D. In below expressions, the assumption is that $P, Q \ll \mu$ for some dominating

measure μ with $p = dP/d\mu$ and $q = dQ/d\mu$.

$$D_{\mathrm{TV}}(P,Q) = \frac{1}{2}\int |p(x) - q(x)| d\mu(x) \tag{17}$$

$$= \int_{x:p(x)\le q(x)} (q(x) - p(x)) d\mu(x) \tag{18}$$

$$= \int_{x:p(x)\ge q(x)} (p(x) - q(x)) d\mu(x). \tag{19}$$

We introduce one more notation. For $\lambda_1, \lambda_2 \in [0,1]$, let $D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2)$ denotes the f-divergence between Bernoulli distributions with $\Pr(1) = \lambda_1$ and λ_2, respectively. That is,

$$D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2) = \lambda_2 f\left(\frac{\lambda_1}{\lambda_2}\right) + (1-\lambda_2) f\left(\frac{1-\lambda_1}{1-\lambda_2}\right). \tag{20}$$

We should note the following facts:

1. By joint convexity of the f-divergence and continuity of f, $D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2)$ is continuous and jointly convex in (λ_1, λ_2). (But $D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2)$ may be extended real-valued)
2. For a fixed λ_1, $D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2)$ attains a global minimum 0 at $\lambda_2 = \lambda_1$. Together with convexity, we derive that $D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2)$ is decreasing in $\lambda_2 \in [0, \lambda_1]$ and increasing in $\lambda_2 \in [\lambda_1, 1]$, respectively.

C Generalized Main Theorem and Proof

In this appendix, we present the formal proofs of the main theorems, Theorems 3.1 and 3.3. As mentioned in Section 3.3, we first state the generalized theorem with its proof, and later we show how this generalized theorem includes the main theorems as special cases.

C.1 Statement of the generalized main theorem

First, let us define the general setup we consider. Let a (general) sample space $\mathcal{X}$ be given. For a positive measure μ on $\mathcal{X}$ such that $\mu(\mathcal{X}) < \infty$ and $c_2 > c_1 \ge 0$, let us define

$$\tilde{\mathcal{P}}_{c_1,c_2,\mu} := \{P \in \mathcal{P}(\mathcal{X}) : P \ll \mu, \quad c_1 \le dP/d\mu \le c_2 \quad \mu\text{-a.e.}\}. \tag{21}$$

We generally consider the case that $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,\mu}$ for some c_1, c_2, μ. For example, for the setup of Section 3.2, we have $\tilde{\mathcal{P}}_{c_1,c_2,h} = \tilde{\mathcal{P}}_{c_1,c_2,\mu}$, where μ is a positive measure with $\mu \ll m$ and $d\mu/dm = h$. For the setup of Section 3.1, we have $\mathcal{P}([k]) = \tilde{\mathcal{P}}_{0,1,\#}$. By the same reason as in Section 3.2, we may impose the following normalization condition on c_1, c_2, μ, ϵ:

$$\mu(\mathcal{X}) = 1, \quad c_1 < 1 < c_2, \quad c_2 > c_1 e^{\epsilon}. \tag{22}$$

Note that $\mu(\mathcal{X}) = 1$ means that μ is a probability measure, that is $\mu \in \mathcal{P}(\mathcal{X})$. Also, note that for $\mathcal{X} = [k]$, we can write in normalized form as $\mathcal{P}([k]) = \tilde{\mathcal{P}}_{0,k,\mu_k}$, where $\mu_k = \frac{1}{k}\#$ is the uniform distribution on $[k]$.

First, let us define the proposed mechanism, together with the related constants as the same as Theorem 3.3. For the ease of proof, we introduce an additional constant α over Theorem 3.3.

Definition C.1. Let c_1, c_2, μ, ϵ satisfying the normalization condition (22) be given, and let $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,\mu}$. First, define the following constants determined by c_1, c_2, ϵ:

$$\alpha = \frac{1-c_1}{c_2 - c_1}, \tag{23}$$

$$b = \frac{c_2 - c_1}{(e^{\epsilon}-1)(1-c_1) + c_2 - c_1} = \frac{1}{\alpha e^{\epsilon} + 1 - \alpha}, \tag{24}$$

$$r_1 = \frac{c_1}{b} = \left(\frac{(e^{\epsilon}-1)(1-c_1) + c_2 - c_1}{c_2 - c_1}\right) c_1 = c_1(\alpha e^{\epsilon} + 1 - \alpha), \tag{25}$$

$$r_2 = \frac{c_2}{be^{\epsilon}} = \left(\frac{(e^{\epsilon}-1)(1-c_1) + c_2 - c_1}{c_2 - c_1}\right) \frac{c_2}{e^{\epsilon}} = \frac{c_2}{e^{\epsilon}}(\alpha e^{\epsilon} + 1 - \alpha). \tag{26}$$

https://doi.org/10.52202/079017-0329

Also, define a mechanism $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon} \in \mathcal{Q}_{\mathcal{X},\tilde{\mathcal{P}},\epsilon}$ as follows:

For each $P \in \tilde{\mathcal{P}}$, $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}(P) =: Q$ *is defined as a probability measure such that* $Q \ll \mu$ *and*

$$\frac{dQ}{d\mu}(x) = \mathrm{clip}\left(\frac{1}{r_P}\frac{dP}{d\mu}(x); b, be^\epsilon\right), \tag{27}$$

where $r_P > 0$ *is a constant depending on* P *so that* $\int \frac{dQ}{d\mu} d\mu(x) = 1$. Furthermore, let $\mathcal{M}_{c_1,c_2,\mu,\epsilon} = \tilde{\mathcal{P}}_{b,be^\epsilon,\mu}$, so that $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}(P) \in \mathcal{M}_{c_1,c_2,\mu,\epsilon}$ for every $P \in \tilde{\mathcal{P}}$.

We should note that $1 = c_2\alpha + c_1(1-\alpha) = be^\epsilon\alpha + b(1-\alpha)$. Also, $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$ clearly satisfies ϵ-LDP.

By the same reason as in Sections 3.1 and 3.2, the values of r_P may not be unique, but the mechanism $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$ does not depend on the choice of r_P. Furthermore, we show that the value of $\int \mathrm{clip}\left(\frac{1}{r}\frac{dP}{d\mu}(x); b, be^\epsilon\right) d\mu(x)$ at $r = r_1$ and $r = r_2$ are at least and at most 1, respectively.

Proposition C.2. *Let* $c_1, c_2, \mu, \epsilon, \tilde{\mathcal{P}}, \alpha, b, r_1, r_2$ *be as in Definition C.1. Then, for any* $P \in \tilde{\mathcal{P}}$, *we have*

$$\int \mathrm{clip}\left(\frac{1}{r_1}\frac{dP}{d\mu}(x); b, be^\epsilon\right) d\mu(x) \geq 1 \geq \int \mathrm{clip}\left(\frac{1}{r_2}\frac{dP}{d\mu}(x); b, be^\epsilon\right) d\mu(x), \tag{28}$$

where, in the case of $r_1 = 0$, *we define*

$$\mathrm{clip}\left(\frac{1}{r_1}\frac{dP}{d\mu}(x); b, be^\epsilon\right) = \begin{cases} b, & \text{if } \frac{dP}{d\mu}(x) = 0 \\ be^\epsilon, & \text{otherwise} \end{cases}. \tag{29}$$

Furthermore, if $\int \mathrm{clip}\left(\frac{1}{r_1}\frac{dP}{d\mu}(x); b, be^\epsilon\right) d\mu(x) = 1$, *then* $\int \mathrm{clip}\left(\frac{1}{r}\frac{dP}{d\mu}(x); b, be^\epsilon\right) d\mu(x) = 1$ *for every* $r \in [r_1, r_2]$.

This proposition, together with the fact that $r \mapsto \int \mathrm{clip}\left(\frac{1}{r}\frac{dP}{d\mu}(x); b, be^\epsilon\right) d\mu(x)$ is continuous and monotone decreasing, implies that r_P can be chosen such that $r_1 < r_P \leq r_2$, as stated in Theorem 3.3. (Again, the continuity is from $\mu(\mathcal{X}) < \infty$ and the dominated convergence theorem)

Also, we should remark that $0 < \alpha < 1$ and $c_1 < b < 1 < be^\epsilon < c_2$. The first one easily follows from $c_1 < 1 < c_2$ and the definition of α. For the second one, first we have $1 < \alpha e^\epsilon + 1 - \alpha < e^\epsilon$, as $\alpha e^\epsilon + 1 - \alpha$ is a propoer convex combination of 1 and ϵ. This directly implies that $b < 1 < be^\epsilon$. Next, by calculations, we can observe that

$$c_1((e^\epsilon - 1)(1 - c_1) + c_2 - c_1) - (c_2 - c_1) = (1 - c_1)(c_1e^\epsilon - c_2) < 0, \tag{30}$$

which implies $c_1 < b$, and

$$c_2((e^\epsilon - 1)(1 - c_1) + c_2 - c_1) - e^\epsilon(c_2 - c_1) = (c_2 - 1)(c_2 - e^\epsilon c_1) > 0, \tag{31}$$

which implies $be^\epsilon < c_2$. Especially, these inequalities imply that $0 \leq r_1 < 1 < r_2$.

Now, we show that under a mild 'decomposability' condition, the proposed mechanism $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$ is universally optimal under any f-divergences for $(\mathcal{X}, \tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,\mu}, \epsilon)$. After that, we show that such a mild condition holds for the setups of both of the main theorems, which finishes the proofs of the main theorems.

First, let us state the 'decomposability' condition. This condition is a formal definition of the concept of "packing of u copies of $\mathcal{X}$ by t subsets $A_1, \cdots, A_t$", which is briefly mentioned in Section 3.3,

Definition C.3. Let $\alpha \in (0,1)$ and $t, u \in \mathbb{N}$, $t > u$. We say that a probability measure $\mu \in \mathcal{P}(\mathcal{X})$ is (α, t, u)**-decomposable** if there exist t subsets $A_1, A_2, \cdots, A_t \subset \mathcal{X}$ such that $\mu(A_i) = \alpha$ for all $i \in [t]$, and for every $x \in \mathcal{X}$, we have $|\{i \in [t] : x \in A_i\}| \leq u$.

We say that $\mu \in \mathcal{P}(\mathcal{X})$ is α**-decomposable** if for any $\delta > 0$, there exists $t, u \in \mathbb{N}$, $t > u$, such that $\alpha \leq u/t < \alpha + \delta$, and μ is (α, t, u)-decomposable.

We remark that $(\alpha, t, 1)$-deomposability means that there are t disjoint subsets $B_1, B_2, \cdots, B_t$ such that $\mu(B_i) = \alpha$ for each $i \in [t]$. Also, if α is a rational number with $\alpha = u/t$, $u, t \in \mathbb{N}$, and μ is (α, t, u)-decomposable, then μ is α-decomposable.

Then, we state the generalized theorem.

Theorem C.4. *Let $c_1, c_2, \mu, \epsilon, \tilde{\mathcal{P}}, \alpha, b, r_1, r_2$ be as in Definition C.1. If μ is α-decomposable, then for any f-divergences D_f, we have*

$$\mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f) = \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1), \tag{32}$$

and furthermore, the mechanism $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$ as in Definition C.1 is optimal for $(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon)$ under any f-divergences D_f.

As guided in Section 3.3, the proof of this theorem is broken into two parts, the achievability part and the converse part.

Proposition C.5 (Achievability part)**.** *Let $c_1, c_2, \mu, \epsilon, \tilde{\mathcal{P}}, \alpha, b, r_1, r_2$ be as in Definition C.1. Then, for any f-divergences D_f, we have*

$$R_f(\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}) \le \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1). \tag{33}$$

(Here, we do not need to assume that μ is α-decomposable)

Proposition C.6 (Converse part)**.** *Let $c_1, c_2, \mu, \epsilon, \tilde{\mathcal{P}}, \alpha, b, r_1, r_2$ be as in Definition C.1, and suppose that μ is α-decomposable. Then for any f-divergences D_f and for any ϵ-LDP mechanism $\mathbf{Q} \in \mathcal{Q}_{\mathcal{X},\tilde{\mathcal{P}},\epsilon}$, we have*

$$R_f(\mathbf{Q}) \ge \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1). \tag{34}$$

The remaining of this appendix is organized as follows. We first present the proofs of Propositions C.5 and C.6 in Appendices C.2 and C.3, in which we grant Proposition C.2 and some intermediate lemmas. After that, in Appendix C.4, we show that Theorem C.4 contains main theorems, Theorems 3.1 and 3.3, as special cases. Finally, Appendix C.5 presents the proof of Proposition C.2, and Appendices C.6 and C.7 prove intermediate lemmas.

C.2 Proof of achievability part (Proposition C.5)

As mentioned in Section 3.3, the proof of the achievability part consists of two steps: first presenting an alternative mechanism and then proving that $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$ performs the f-divergence projection for any general f-divergence.

C.2.1 An alternative mechanism

Let $\mathbf{Q}^\dagger_{c_1,c_2,\mu,\epsilon} \in \mathcal{Q}_{\mathcal{X},\tilde{\mathcal{P}},\epsilon}$ be a mechanism defined as follows:

$$\mathbf{Q}^\dagger_{c_1,c_2,\mu,\epsilon}(P) = \gamma P + (1-\gamma)\mu, \tag{35}$$

where

$$\gamma = \frac{e^\epsilon - 1}{(e^\epsilon - 1)(1-c_1) + c_2 - c_1}. \tag{36}$$

Since $c_2 > e^\epsilon c_1$, it follows that $0 < \gamma < 1$. Also, a direct computation shows that

$$b = \gamma c_1 + (1-\gamma), \tag{37}$$

$$be^\epsilon = \gamma c_2 + (1-\gamma). \tag{38}$$

Notice that since $c_1 \le \frac{dP}{d\mu} \le c_2$ and $\frac{d\mathbf{Q}^\dagger_{c_1,c_2,\mu,\epsilon}(P)}{d\mu} = \gamma \frac{dP}{d\mu} + (1-\gamma)$, we have

$$\frac{d\mathbf{Q}^\dagger_{c_1,c_2,\mu,\epsilon}(P)}{d\mu}(x) \ge \gamma c_1 + (1-\gamma) = b, \tag{39}$$

$$\frac{d\mathbf{Q}^\dagger_{c_1,c_2,\mu,\epsilon}(P)}{d\mu}(x) \le \gamma c_2 + (1-\gamma) = be^\epsilon. \tag{40}$$

https://doi.org/10.52202/079017-0329

Hence, $\mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}(P) \in \mathcal{M}_{c_1,c_2,\mu,\epsilon}$ for every $P \in \tilde{\mathcal{P}}$, implying that $\mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}$ also satisfies ϵ-LDP. Now, we show that

$$R_f(\mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}) \leq \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1). \tag{41}$$

To this end, we need to show that for each $P \in \tilde{\mathcal{P}}$, we have

$$D_f\left(P \middle\| \mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}(P)\right) \leq \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1). \tag{42}$$

Fix $P \in \tilde{\mathcal{P}}$. Let $p = dP/d\mu$ and $q = d\mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}(P)/d\mu$. First, we claim that

$$r_1 \leq \frac{p(x)}{q(x)} \leq r_2 \tag{43}$$

for μ-almost every $x \in \mathcal{X}$. We have

$$\frac{p(x)}{q(x)} = \frac{p(x)}{\gamma p(x) + (1-\gamma)} \tag{44}$$

$$= \frac{1}{\gamma}\left(1 - \frac{1-\gamma}{\gamma p(x) + (1-\gamma)}\right), \tag{45}$$

which is increasing in $p(x) \geq 0$. Since $c_1 \leq p(x) \leq c_2$, we have

$$\frac{c_1}{\gamma c_1 + (1-\gamma)} \leq \frac{p(x)}{q(x)} \leq \frac{c_2}{\gamma c_2 + (1-\gamma)} \tag{46}$$

for μ-almost every $x \in \mathcal{X}$. From (37), (38), and Definition C.1, we conclude that

$$r_1 \leq \frac{p(x)}{q(x)} \leq r_2, \tag{47}$$

which proves the claim. For the remaining of the proof, we need the following lemma.

Lemma C.7 ([59], Theorem 2.1)**.** *Let $P, Q \in \mathcal{P}(\mathcal{X})$. Suppose that $P, Q \ll \mu$ for some reference measure μ on $\mathcal{X}$, and there exist $r_1, r_2 \in \mathbb{R}$ with $0 \leq r_1 < 1 < r_2$ such that the densities $p = \frac{dP}{d\mu}, q = \frac{dQ}{d\mu}$ satisfy $q(x) > 0$ and $r_1 \leq \frac{p(x)}{q(x)} \leq r_2$ for μ-almost every $x \in \mathcal{X}$. Then for any f-divergence D_f, we have*

$$D_f\left(P \| Q\right) \leq \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1). \tag{48}$$

This lemma directly implies (42), and consequently (41).

C.2.2 f-divergence projection of the proposed mechanism

Next, we prove that $\mathbf{Q}^{*}_{c_1,c_2,\mu,\epsilon}(P)$ is the projection of P onto $\mathcal{M}_{c_1,c_2,\mu,\epsilon}$ for *every f-divergence*. Proposition 3.4 can be stated in a more general way as follows.

Proposition C.8. *For any $P \in \tilde{\mathcal{P}}$ and any f-divergences D_f, we have*

$$D_f\left(P \middle\| \mathbf{Q}^{*}_{c_1,c_2,\mu,\epsilon}(P)\right) = \inf_{Q \in \mathcal{M}_{c_1,c_2,\mu,\epsilon}} D_f\left(P \| Q\right). \tag{49}$$

This proposition implies that

$$D_f\left(P \middle\| \mathbf{Q}^{*}_{c_1,c_2,\mu,\epsilon}(P)\right) \leq D_f\left(P \middle\| \mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}(P)\right), \tag{50}$$

for every $P \in \tilde{\mathcal{P}}$, and hence

$$R_f(\mathbf{Q}^{*}_{c_1,c_2,\mu,\epsilon}) \leq R_f(\mathbf{Q}^{\dagger}_{c_1,c_2,\mu,\epsilon}) \leq \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1), \tag{51}$$

which completes the proof of the achievability part.

Next, we prove Proposition C.8. Fix $P \in \tilde{\mathcal{P}}$ and $p = dP/d\mu$. If $f(0) = \infty$ and $\mu(\{x : p(x) = 0\}) > 0$, then $D_f(P\|Q) = \infty$ for every $Q \in \mathcal{M}_{c_1,c_2,\mu,\epsilon}$, thus the proposition holds trivially. Consequently, we assume either $f(0) < \infty$ or $\mu(\{x : p(x) = 0\}) = 0$.

The optimization problem $\inf_{Q \in \mathcal{M}_{c_1,c_2,\mu,\epsilon}} D_f(P\|Q)$ can be cast as the following

$$\inf_{q:\mathcal{X}\to(0,\infty)} \int q(x) f\left(\frac{p(x)}{q(x)}\right) d\mu(x) \tag{52}$$

$$\text{such that} \quad q(x) \geq b, \quad \forall x, \tag{53}$$

$$q(x) \leq be^{\epsilon}, \quad \forall x, \tag{54}$$

$$\int q(x) d\mu(x) = 1, \tag{55}$$

where $q = dQ/d\mu$. Our goal is to show that $q^* = d\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}(P)/d\mu$ is an optimal solution of the above optimization problem.

Motivation for the optimality proof. To provide the motivation for the proof of the optimality of q^* for the above optimization problem, we first consider the following analogous finite-dimensional optimization problem

$$\inf_{q\in(0,\infty)^n} \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{56}$$

$$\text{such that} \quad q_i \geq b, \tag{57}$$

$$q_i \leq be^{\epsilon}, \tag{58}$$

$$\sum_{i=1}^{n} q_i = 1. \tag{59}$$

for convex function $f : (0,\infty) \to \mathbb{R}$ and $p_i > 0$. The convexity of $f^\star(x) = xf(1/x)$ (see Appendix B) implies that the above is a convex optimization problem. For simplicity, we assume that $f^\star$ is differentiable. However, note that this assumption is only for simplifying the motivation for the proof, and the result holds for general f or $f^\star$.

We formulate the Lagrangian for the above optimization as

$$\mathcal{L}(q,\phi,\psi,\nu) = \sum_{i=1}^{n} (p_i f^\star(q_i/p_i) + \phi_i(b - q_i) + \psi_i(q_i - be^{\epsilon})) + \nu\left(1 - \sum_{i=1}^{n} q_i\right) \tag{60}$$

with dual variables $\phi, \psi \in [0,\infty)^n$ and $\nu \in \mathbb{R}$.

The Karush-Kuhn-Tucker (KKT) condition yields:

$$(f^\star)'(q_i/p_i) - \phi_i + \psi_i - \nu = 0, \quad \forall i, \tag{61}$$

$$(57), (58), (59), \tag{62}$$

$$\phi_i, \psi_i \geq 0, \quad \forall i, \tag{63}$$

$$\phi_i(q_i - b) = \psi_i(be^{\epsilon} - q_i) = 0, \quad \forall i. \tag{64}$$

Now, suppose that there is a feasible point $q^* \in (0,\infty)^n$ satisfying (57), (58), and (59) such that $q^* = \mathrm{clip}(p_i/r; b, be^{\epsilon})$ for some $r > 0$. We show that q^* satisfies the KKT condition for some feasible dual variables (ϕ,ψ,ν).

For (q^*,ϕ,ψ,ν) to satisfy the KKT condition, the following should hold:

- If $b < p_i/r < be^{\epsilon}$, then we have $q_i^* = p_i/r$. From (61) and (64), we must have $\phi_i = \psi_i = 0$ and $\nu = (f^\star)'(1/r)$.
- If $p_i/r \leq b$, then $q_i^* = b$. Then, $\psi_i = 0$ and $\phi_i = (f^\star)'(b/p_i) - \nu$.
- If $p_i/r \geq be^{\epsilon}$, then $q_i^* = be^{\epsilon}$. Then $\phi_i = 0$ and $\psi_i = \nu - (f^\star)'(be^{\epsilon}/p_i)$.

 https://doi.org/10.52202/079017-0329

Now, since $f^\star$ is convex, $(f^\star)'$ is monotonically increasing. Hence, the following should be satisfied:

- If $p_i/r \le b$, then $(f^\star)'(b/p_i) - (f^\star)'(1/r) \ge 0$, and
- If $p_i/r \ge be^\epsilon$, then $(f^\star)'(1/r) - (f^\star)'(be^\epsilon/p_i) \ge 0$.

It can thus be verified that (q^*, ϕ, ψ, ν) satisfies the KKT condition with

- $\nu = (f^\star)'(1/r)$,
- $\phi_i = \begin{cases} (f^\star)'(b/p_i) - (f^\star)'(1/r), & \text{if } p_i/r \le b, \\ 0, & \text{otherwise,} \end{cases}$
- $\psi_i = \begin{cases} (f^\star)'(1/r) - (f^\star)'(be^\epsilon/p_i), & \text{if } p_i/r \ge be^\epsilon, \\ 0, & \text{otherwise.} \end{cases}$

Optimality proof. Next, we present a proof for the optimality of q^* for the optimization problem

$$\inf_{q:\mathcal{X}\to(0,\infty)} \int q(x) f\left(\frac{p(x)}{q(x)}\right) d\mu(x) \tag{65}$$

$$\text{such that} \quad q(x) \ge b, \quad \forall x, \tag{66}$$

$$q(x) \le be^\epsilon, \quad \forall x, \tag{67}$$

$$\int q(x) d\mu(x) = 1. \tag{68}$$

Here, note that we do not assume f is differentiable. To this goal, we first review the following basic facts about a general convex function $f : (0, \infty) \to \mathbb{R}$, which *may not be differentiable* [64, 65]:

- The left derivative $f'_-(x) := \lim_{h\to 0-} \frac{f(x+h)-f(x)}{h}$ and the right derivative $f'_+(x) := \lim_{h\to 0+} \frac{f(x+h)-f(x)}{h}$ exist and finite for every $x \in (0, \infty)$, regardless of whether f is differentiable or not.
- For every $0 < x < y$, we have $f'_-(x) \le f'_+(x) \le f'_-(y) \le f'_+(y)$.
- For every $x, y \in (0, \infty)$ and any $g \in [f'_-(x), f'_+(x)]$, we have

$$f(y) \ge f(x) + g(y - x). \tag{69}$$

  By continuous extension, this holds for $y = 0$ also. That is, $f(0) \ge f(x) - gx$ for every $x \in (0, \infty)$ and $g \in [f'_-(x), f'_+(x)]$.

Let $q : \mathcal{X} \to (0, \infty)$ be any feasible function satisfying (66) - (68). Recall that $f^\star(x) := xf(1/x)$ is convex, and we can express $q(x)f(p(x)/q(x)) = p(x)f^\star(q(x)/p(x))$ whenever $p(x) \ne 0$. Also, whenever $p(x) \ne 0$, we bound $f^\star(q(x)/p(x))$ by the linear approximation of $f^\star$ at $q^*(x)/p(x)$ using (69), as follows:

$$f^\star\left(\frac{q(x)}{p(x)}\right) \ge f^\star\left(\frac{q^*(x)}{p(x)}\right) + (f^\star)'_+\left(\frac{q^*(x)}{p(x)}\right)\left[\frac{q(x)}{p(x)} - \frac{q^*(x)}{p(x)}\right]. \tag{70}$$

Hence, we have

$$q(x)f\left(\frac{p(x)}{q(x)}\right) = p(x)f^\star\left(\frac{q(x)}{p(x)}\right) \ge q(x)\zeta(x) + \xi(x), \tag{71}$$

where

$$\zeta(x) = (f^\star)'_+\left(\frac{q^*(x)}{p(x)}\right), \tag{72}$$

$$\xi(x) = p(x)f^\star\left(\frac{q^*(x)}{p(x)}\right) - q^*(x)(f^\star)'_+\left(\frac{q^*(x)}{p(x)}\right). \tag{73}$$

Also, whenever $p(x) = 0$, we have $q(x)f(p(x)/q(x)) = q(x)p(0)$. Hence, we set $\zeta(x) = p(0)$ and $\xi(x) = 0$ when $p(x) = 0$, so that

$$q(x)f\left(\frac{p(x)}{q(x)}\right) \geq q(x)\zeta(x) + \xi(x) \tag{74}$$

holds for all $x \in \mathcal{X}$. We note that $\zeta(x)$ and $\xi(x)$ does not depend on the choice of $q(x)$. Also, the equality holds if $q(x) = q^*(x)$.

Next, as an analogous to ν in the above motivation for the proof, let $\nu = (f^\star)'_+(1/r_P)$. Since $\int q(x)d\mu(x) = 1$, we can write

$$\int q(x)f\left(\frac{p(x)}{q(x)}\right)d\mu(x) = \int q(x)\left(f\left(\frac{p(x)}{q(x)}\right) - \nu\right)d\mu(x) + \nu \tag{75}$$

$$\geq \int \left[q(x)(\zeta(x) - \nu) + \xi(x)\right]d\mu(x) + \nu. \tag{76}$$

Now, we define the following sets which form a partition of $\mathcal{X}$:

$$L = \left\{x \in \mathcal{X} : \frac{1}{r_P}p(x) < b\right\}, \tag{77}$$

$$M = \left\{x \in \mathcal{X} : b \leq \frac{1}{r_P}p(x) \leq be^\epsilon\right\}, \tag{78}$$

$$U = \left\{x \in \mathcal{X} : \frac{1}{r_P}p(x) > be^\epsilon\right\}. \tag{79}$$

For each case of $x \in L, M, U$, we have $q^*(x) = b, \frac{1}{r_P}p(x), be^\epsilon$, respectively. We observe the following:

- If $x \in M$, then $\zeta(x) = \nu$ clearly.
- If $x \in U$, then $q^*(x)/p(x) < 1/r_P$. Since $(f^*)'_+$ is monotone increasing, we have $\zeta(x) \leq \nu$.
- If $x \in L$ and $p(x) \neq 0$, then $q^*(x)/p(x) > 1/r_P$. Again, since $(f^*)'_+$ is monotone increasing, we have $\zeta(x) \geq \nu$.
- Finally, if $p(x) = 0$ (which implies $x \in L$ also), then from the definition $f^\star(t) = tf(1/t)$, we have $(f^\star)'_+(t) = f(1/t) - \frac{1}{t}f'_-(1/t)$. From (69) with $y = 0$, we have $\nu = (f^\star)'_+(1/r_P) = f(r_P) - r_P f'_-(r_P) \leq f(0) = \zeta(x)$.
  Hence, $\zeta(x) \geq \nu$ for every $x \in L$.

Therefore, since $b \leq q(x) \leq be^\epsilon$, we can write

$$\int q(x)f\left(\frac{p(x)}{q(x)}\right)d\mu(x) \geq \int \left[q(x)(\zeta(x) - \nu) + \xi(x)\right]d\mu(x) + \nu \tag{80}$$

$$\geq [b(\zeta(x) - \nu)\mathbb{1}_L(x) + be^\epsilon(\zeta(x) - \nu)\mathbb{1}_U(x) + \xi(x)]d\mu(x) + \nu. \tag{81}$$

The last expression does not depend on the choice of $q(x)$. Also, we observe that all inequalities become equality if $q(x) = q^*(x)$. This completes the proof for the optimality of q^*, and hence completes the proof of the achievability part. □

C.3 Proof of converse part (Proposition C.6)

Let $\mathbf{Q} \in \mathcal{Q}_{\mathcal{X},\tilde{\mathcal{P}},\epsilon}$ be given. Let $\mathcal{A} = \{A \subset \mathcal{X} : \mu(A) = \alpha\}$. For each $A \in \mathcal{A}$, let $p_A(x) = \begin{cases} c_2 & \text{if } x \in A \\ c_1 & \text{if } x \in \mathcal{X}\backslash A \end{cases}$. Since $c_2\alpha + c_1(1-\alpha) = 1$, we have $\int p_A(x)d\mu(x) = 1$. Hence for each $A \in \mathcal{A}$, we can define a probability measure $P_A \in \tilde{\mathcal{P}}$ by $\frac{dP_A}{d\mu} = p_A$. Also, note that $P_A(A) = c_2\alpha$.

For each $A \in \mathcal{A}$, let $\beta_A = \mathbf{Q}(A|P_A)$. Then, the push-forward measures of P_A and $\mathbf{Q}(P_A)$ by the indicator function $\mathbb{1}_A$ are Bernoulli distributions with $\Pr(1) = c_2\alpha$ and β_A, respectively. By the data processing inequality (Theorem B.2), we have

$$D_f\left(P_A \| \mathbf{Q}(P_A)\right) \geq D_f^{\mathrm{B}}\left(c_2\alpha \| \beta_A\right). \tag{82}$$

 https://doi.org/10.52202/079017-0329

The main lemma to proceed is the following.

Lemma C.9. *Let $\mathcal{X}$ be a sample space. Let $t, u \in \mathbb{N}$, $t > u$. Let $A_1, A_2, \cdots, A_t \subset \mathcal{X}$ be subsets such that for each $x \in \mathcal{X}$, we have $|\{i \in [t] : x \in A_i\}| \leq u$. Then for any t probability measures $Q_1, \cdots, Q_t \in \mathcal{P}(\mathcal{X})$ satisfying that $Q_i(A) \leq e^{\epsilon} Q_j(A)$ for all $i, j \in [t]$ and $A \subset \mathcal{X}$, we have*

$$\min_{i \in [t]} Q_i(A_i) \leq \frac{(u/t)e^{\epsilon}}{(u/t)e^{\epsilon} + 1 - (u/t)}. \tag{83}$$

Now, by the assumption that μ is α-decomposable, there exist sequences $\{t_j\}_{j=1}^{\infty}, \{u_j\}_{j=1}^{\infty} \subset \mathbb{N}$ of positive integers such that $t_j > u_j$, $\alpha \leq u_j/t_j$, $\lim_{j\to\infty} u_j/t_j = \alpha$, and for each $j \in \mathbb{N}$, there exist t_j subsets $A_{j,1}, A_{j,2}, \cdots, A_{j,t_j} \subset \mathcal{X}$ such that $\mu(A_{j,i}) = \alpha$ for all $i \in [t_j]$, and for every $x \in \mathcal{X}$, we have $\left|\left\{i \in [t_j] : x \in A_{j,t_j}\right\}\right| \leq u_j$. By applying Lemma C.9 to $Q_i = \mathbf{Q}(P_{A_{j,i}})$, we obtain $\min_{i\in[t_j]} \beta_{A_{i,j}} \leq \frac{(u_j/t_j)e^{\epsilon}}{(u_j/t_j)e^{\epsilon}+1-(u_j/t_j)}$. This implies that $\inf_{A\in\mathcal{A}} \beta_A \leq \frac{(u_j/t_j)e^{\epsilon}}{(u_j/t_j)e^{\epsilon}+1-(u_j/t_j)}$ for all $j \in \mathbb{N}$, and by taking the limit $j \to \infty$, we have $\inf_{A\in\mathcal{A}} \beta_A \leq \frac{\alpha e^{\epsilon}}{\alpha e^{\epsilon}+1-\alpha} = be^{\epsilon}\alpha$.

Since $be^{\epsilon} < c_2$, we have $be^{\epsilon}\alpha < c_2\alpha$. By continuity of D_f^{B} and the fact that $D_f^{\mathrm{B}}(\lambda_1 \| \lambda_2)$ is decreasing in $\lambda_2 \in [0, \lambda_1]$, we have

$$\sup_{A\in\mathcal{A}} D_f(P_A \| \mathbf{Q}(P_A)) \geq \sup_{A\in\mathcal{A}} D_f^{\mathrm{B}}(c_2\alpha \| \beta_A) \geq D_f^{\mathrm{B}}(c_2\alpha \| be^{\epsilon}\alpha). \tag{84}$$

It follows that

$$R_f(\mathbf{Q}) = \sup_{P\in\bar{\mathcal{P}}} D_f(P \| \mathbf{Q}(P)) \geq \sup_{A\in\mathcal{A}} D_f(P_A \| \mathbf{Q}(P_A)) \geq D_f^{\mathrm{B}}(c_2\alpha \| be^{\epsilon}\alpha). \tag{85}$$

Furthermore, we have

$$\begin{aligned} D_f^{\mathrm{B}}(c_2\alpha \| be^{\epsilon}\alpha) &= be^{\epsilon}\alpha f\left(\frac{c_2\alpha}{be^{\epsilon}\alpha}\right) + (1 - b\alpha e^{\epsilon}) f\left(\frac{1-c_2\alpha}{1-be^{\epsilon}\alpha}\right) && (86)\\ &= be^{\epsilon}\alpha f\left(\frac{c_2\alpha}{be^{\epsilon}\alpha}\right) + b(1-\alpha) f\left(\frac{c_1(1-\alpha)}{b(1-\alpha)}\right) && (87)\\ &= be^{\epsilon}\alpha f(r_2) + b(1-\alpha) f(r_1), && (88) \end{aligned}$$

and we can derive $be^{\epsilon}\alpha = \frac{1-r_1}{r_2-r_1}$ and $b(1-\alpha) = \frac{r_2-1}{r_2-r_1}$, as follows. From (30), (31) and the definition of r_1, r_2, we have

$$\begin{aligned} 1 - r_1 &= (1-c_1)\frac{c_2 - c_1 e^{\epsilon}}{c_2 - c_1}, && (89)\\ r_2 - 1 &= \frac{c_2 - 1}{e^{\epsilon}} \frac{c_2 - e^{\epsilon} c_1}{c_2 - c_1}. && (90) \end{aligned}$$

From this, we have

$$\begin{aligned} \frac{1-r_1}{r_2-r_1} &= \frac{1-r_1}{(r_2-1)+(1-r_1)} && (91)\\ &= \frac{(1-c_1)e^{\epsilon}}{(1-c_1)e^{\epsilon} + (c_2-1)} && (92)\\ &= \frac{(1-c_1)e^{\epsilon}}{(e^{\epsilon}-1)(1-c_1) + c_2 - c_1} && (93)\\ &= b \times e^{\epsilon} \frac{1-c_1}{c_2-c_1} && (94)\\ &= be^{\epsilon}\alpha. && (95) \end{aligned}$$

Also, from above and $1 = be^{\epsilon}\alpha + b(1-\alpha)$, we have

$$\frac{r_2-1}{r_2-r_1} = 1 - \frac{1-r_1}{r_2-r_1} = 1 - be^{\epsilon}\alpha = b(1-\alpha). \tag{96}$$

This ends the proof of the converse part. □

C.4 Deduction to main theorems

In this subsection, we show that Theorem C.4 contains main theorems, Theorems 3.1 and 3.3, as special cases.

C.4.1 Deduction to Theorem 3.1

Recall that in this setup, $\mathcal{X} = [k]$, $\tilde{\mathcal{P}} = \mathcal{P}([k]) = \tilde{\mathcal{P}}_{0,k,\mu_k}$, where μ_k is the uniform distribution on $[k]$. That is, $(c_1, c_2) = (0, k)$. The values of the constants α, b, r_1, r_2 are

$$\alpha = 1/k, \tag{97}$$

$$b = \frac{k}{e^\epsilon + k - 1}, \tag{98}$$

$$r_1 = 0, \tag{99}$$

$$r_2 = \frac{e^\epsilon + k - 1}{e^\epsilon}. \tag{100}$$

We can easily observe that μ_k is $(\alpha, k, 1)$-decomposable, because the sets $\{i\}$, $i \in [k]$, are disjoint and $\mu_k(\{i\}) = \frac{1}{k}$. It follows that μ_k is α-decomposable. Hence, we can apply Theorem C.4. By a direct calculation of $\mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f) = \frac{1-r_1}{r_2-r_1} f(r_2) + \frac{r_2-1}{r_2-r_1} f(r_1)$, we can derive the formula of $\mathcal{R}([k], \mathcal{P}([k]), \epsilon, f)$ presented in Theorem 3.1. It remains to show that the mechanism $\mathbf{Q}^*_{k,\epsilon}$ presented in Theorem 3.1 is the same as $\mathbf{Q}^*_{0,k,\mu_k,\epsilon}$, and show the claim about the range of r_P. Recall that

$$\mathbf{Q}^*_{k,\epsilon}(x|P) = \max\left(\frac{1}{r_P} P(x), \frac{1}{e^\epsilon + k - 1}\right) \quad \forall x \in [k], P \in \mathcal{P}([k]), \tag{101}$$

and we claim that r_P can be chosen such that $1 \le r_P \le (e^\epsilon + k - 1)/e^\epsilon$.

First, as explained in Section 3.1, we can alternatively write

$$\mathbf{Q}^*_{k,\epsilon}(x|P) = \mathrm{clip}\left(\frac{1}{r_P} P(x); \frac{1}{e^\epsilon + k - 1}, \frac{e^\epsilon}{e^\epsilon + k - 1}\right). \tag{102}$$

Since $P(x) = \frac{1}{k} \frac{dP}{d\mu_k}(x)$ for each $x \in [k]$ and $P \in \mathcal{P}([k])$, (102) can be written as

$$\frac{1}{k} \frac{d\mathbf{Q}^*_{k,\epsilon}(P)}{d\mu_k}(x) = \mathrm{clip}\left(\frac{1}{r_P} \frac{1}{k} \frac{dP}{d\mu_k}(x); \frac{1}{e^\epsilon + k - 1}, \frac{e^\epsilon}{e^\epsilon + k - 1}\right) \tag{103}$$

$$= \frac{1}{k} \mathrm{clip}\left(\frac{1}{r_P} \frac{dP}{d\mu_k}(x); \frac{k}{e^\epsilon + k - 1}, \frac{ke^\epsilon}{e^\epsilon + k - 1}\right), \tag{104}$$

hence,

$$\frac{d\mathbf{Q}^*_{k,\epsilon}(P)}{d\mu_k}(x) = \mathrm{clip}\left(\frac{1}{r_P} \frac{dP}{d\mu_k}(x); \frac{k}{e^\epsilon + k - 1}, \frac{ke^\epsilon}{e^\epsilon + k - 1}\right) \tag{105}$$

$$= \mathrm{clip}\left(\frac{1}{r_P} \frac{dP}{d\mu_k}(x); b, be^\epsilon\right). \tag{106}$$

Hence, $\mathbf{Q}^*_{k,\epsilon} = \mathbf{Q}^*_{0,k,\mu_k,\epsilon}$.

Second, to prove the claim about the range of r_P, let us fix $P \in \mathcal{P}([k])$. We need to show that

$$\sum_{x=1}^{k} \max\left(P(x), \frac{1}{e^\epsilon + k - 1}\right) \ge 1 \ge \sum_{x=1}^{k} \max\left(\frac{e^\epsilon}{e^\epsilon + k - 1} P(x), \frac{1}{e^\epsilon + k - 1}\right). \tag{107}$$

The left inequality can be easily derived by

$$\sum_{x=1}^{k} \max\left(P(x), \frac{1}{e^\epsilon + k - 1}\right) \ge \sum_{x=1}^{k} P(x) = 1. \tag{108}$$

https://doi.org/10.52202/079017-0329

For the right inequality, we recall that $b = \frac{k}{e^\epsilon+k-1}$. Since $P(x) \le 1$, we have $\frac{e^\epsilon}{e^\epsilon+k-1}P(x) \le \frac{e^\epsilon}{e^\epsilon+k-1}$. Hence, again by using $P(x) = \frac{1}{k}\frac{dP}{d\mu_k}(x)$, we have

$$\max\left(\frac{e^\epsilon}{e^\epsilon+k-1}P(x), \frac{1}{e^\epsilon+k-1}\right) \tag{109}$$

$$=\text{clip}\left(\frac{e^\epsilon}{e^\epsilon+k-1}P(x); \frac{1}{e^\epsilon+k-1}, \frac{e^\epsilon}{e^\epsilon+k-1}\right) \tag{110}$$

$$=\text{clip}\left(\frac{e^\epsilon}{e^\epsilon+k-1}\frac{1}{k}\frac{dP}{d\mu_k}(x); \frac{1}{e^\epsilon+k-1}, \frac{e^\epsilon}{e^\epsilon+k-1}\right) \tag{111}$$

$$= \frac{1}{k}\text{clip}\left(\frac{1}{r_2}\frac{dP}{d\mu_k}(x); b, be^\epsilon\right), \tag{112}$$

and thus

$$\sum_{x=1}^{k}\max\left(\frac{e^\epsilon}{e^\epsilon+k-1}P(x), \frac{1}{e^\epsilon+k-1}\right) \tag{113}$$

$$= \sum_{x=1}^{k}\frac{1}{k}\text{clip}\left(\frac{1}{r_2}\frac{dP}{d\mu_k}(x); b, be^\epsilon\right) \tag{114}$$

$$= \int \text{clip}\left(\frac{1}{r_2}\frac{dP}{d\mu_k}(x); b, be^\epsilon\right) d\mu_k(x). \tag{115}$$

The desired inequality follows from Proposition C.2.

C.4.2 Deduction to Theorem 3.3

Recall that in this setup, $\mathcal{X} = \mathbb{R}^n$, $\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{c_1,c_2,h} = \tilde{\mathcal{P}}_{c_1,c_2,\mu}$, where $\mu \ll m$ and $d\mu/dm = h$. Note that the normalization condition (10) about (c_1, c_2, h, ϵ) implies the normalization condition (22) about $(c_1, c_2, \mu, \epsilon)$. It can be directly observed that $\mathbf{Q}^*_{c_1,c_2,h,\epsilon} = \mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$, because for each $P \in \tilde{\mathcal{P}}$ with corresponding pdf $p(x)$, the chain rule of the Radon-Nikodym derivative shows that $p(x) = \frac{dP}{dm} = \frac{dP}{d\mu}(x)\frac{d\mu}{dm}(x) = \frac{dP}{d\mu}(x)h(x)$. Hence, once we show that μ is α-decomposable, Theorem C.4 and Proposition C.2 directly contain Theorem 3.3 as a special case. Thus, it remains to show μ is α-decomposable.

In fact, we prove a stronger statement that: μ is (α, t, u)-decomposable for any $t, u \in \mathbb{N}$ such that $t > u$ and $\alpha \le u/t$. Then, since the set of rational numbers is dense in $\mathbb{R}$, this also implies that μ is α-decomposable.

To prove this, let us first introduce the following lemma.

Lemma C.10. *Let $\alpha \in (0,1)$ and $t, u \in \mathbb{N}$, $t > u$, $\alpha \le u/t$. If $\mu \in \mathcal{P}(\mathcal{X})$ is $(\alpha/u, t, 1)$-decomposable, then μ is also (α, t, u)-decomposable.*

Proof. In this proof, assume that the sum and subtraction operations performed in subscripts are modulo t operations, with the identification that $0 = t$.

By $(\alpha/u, t, 1)$-decomposability, there are t disjoint subsets $B_1, B_2, \cdots, B_t$ such that $\mu(B_i) = \alpha/u$ for each $i \in [t]$. Using this, for each $i \in [t]$, define A_i as $A_i = \cup_{j=0}^{u-1} B_{i+j}$. As B_i's are disjoint, we have $\mu(A_i) = \sum_{j=0}^{u-1}\mu(B_{i+j}) = u \times (\alpha/u) = \alpha$ for all $i \in [t]$. Also, for each $x \in B_i$, x is contained in exactly u sets among $A_1, \cdots, A_t$, which are $A_i, A_{i-1}, \cdots, A_{i-u+1}$. Furthermore, if $x \notin B_i$ for all $i \in [t]$, then x is contained in none of A_i. Hence $|\{i \in [t] : x \in A_i\}| \le u$ for all $x \in \mathcal{X}$. Thus μ is (α, t, u)-decomposable. □

By this lemma, it suffices to show that for any $t \in \mathbb{N}$ such that $t \ge 2$ and $\alpha \le 1/t$, μ is $(\alpha, t, 1)$-decomposable. As $\mu \ll m$, the map $s \in \mathbb{R} \mapsto \mu((-\infty, s] \times \mathbb{R}^{n-1})$ is continuous, and as $s \to -\infty$ and ∞, we have $\mu((-\infty, s] \to 0$ and 1, respectively. Hence by the intermediate value theorem, for each $i \in [t]$, there exists $s_i \in \mathbb{R}$ such that $\mu((-\infty, s] \times \mathbb{R}^{n-1}) = \alpha i$. Then, setting $A_1 = (-\infty, s_1] \times \mathbb{R}^{n-1}$ and $A_i = (s_{i-1}, s_i] \times \mathbb{R}^{n-1}$ for $i \ge 2$ gives the desired A_i's in the definition of decomposability.

In conclusion, μ is α-decomposable, and hence Theorem 3.3 can be deduced from Theorem C.4.

C.5 Proof of Proposition C.2

Let $P \in \tilde{\mathcal{P}}$ be given. Let $p = \frac{dP}{d\mu}$, and again assume that $c_1 \le p(x) \le c_2$ for *all* $x \in \mathcal{X}$. Similar to the proof of the achievability part, for each $r > 0$, let us define the following sets which form a partition of $\mathcal{X}$:

$$L_r = \left\{ x \in \mathcal{X} : \frac{1}{r} p(x) < b \right\}, \tag{116}$$

$$M_r = \left\{ x \in \mathcal{X} : b \le \frac{1}{r} p(x) \le b e^{\epsilon} \right\}, \tag{117}$$

$$U_r = \left\{ x \in \mathcal{X} : \frac{1}{r} p(x) > b e^{\epsilon} \right\}. \tag{118}$$

For $r = 0$, we let $L_0 = \{x \in \mathcal{X} : p(x) = 0\}$, $U_0 = \{x \in \mathcal{X} : p(x) > 0\}$, $M_0 = \emptyset$. Also, let $q_r(x) = \operatorname{clip}\left(\frac{1}{r} p(x); b, b e^{\epsilon}\right)$. Then, $q_r(x) = b, \frac{1}{r} p(x), b e^{\epsilon}$ for $x \in L_r, M_r, U_r$, respectively.

First, we show that $\int q_{r_1}(x) d\mu(x) \ge 1$. We divide the case of $c_1 = 0$ and $c_1 > 0$.

Suppose first that $c_1 = 0$. Then $r_1 = 0, \alpha = \frac{1}{c_2}$, and $b = \frac{c_2}{e^{\epsilon} + c_2 - 1}$. Observe that

$$1 = \int p(x) d\mu(x) = \int_{U_0} p(x) d\mu(x) \le c_2 \mu(U_0), \tag{119}$$

hence $\mu(U_0) \ge 1/c_2$. It follows that

$$\begin{aligned}
\int q_{r_1}(x) d\mu(x) &= b\mu(L_0) + b e^{\epsilon} \mu(U_0) && (120)\\
&= b(1 - \mu(U_0)) + b e^{\epsilon} \mu(U_0) && (121)\\
&= b(e^{\epsilon} - 1)\mu(U_0) + b && (122)\\
&\ge \frac{b(e^{\epsilon} - 1)}{c_2} + b && (123)\\
&= b \times \frac{e^{\epsilon} + c_2 - 1}{c_2} = 1. && (124)
\end{aligned}$$

Next, suppose that $c_1 > 0$. Then $r_1 > 0$. From $p(x) \ge c_1$, we have $\frac{1}{r_1} p(x) \ge \frac{c_1}{r_1} = b$. Hence $L_{r_1} = \emptyset$. Thus

$$\begin{aligned}
\int q_{r_1}(x) d\mu(x) &= \frac{1}{r_1} \int_{M_{r_1}} p(x) d\mu(x) + b e^{\epsilon} \mu(U_{r_1}) && (125)\\
&= \frac{1}{r_1} \left(\int_{M_{r_1}} p(x) d\mu(x) + c_1 e^{\epsilon} \mu(U_{r_1}) \right). && (126)
\end{aligned}$$

Let $S_1 = \int_{M_{r_1}} p(x) d\mu(x)$ and $T_1 = \mu(U_{r_1})$, so that

$$\int q_{r_1}(x) d\mu(x) = \frac{1}{r_1}(S_1 + c_1 e^{\epsilon} T_1). \tag{127}$$

As $c_1 \le p(x) \le c_2$, we have

$$S_1 = \int_{M_{r_1}} p(x) d\mu(x) \ge c_1 \mu(M_{r_1}) = c_1(1 - \mu(U_{r_1})) = c_1(1 - T_1), \tag{128}$$

and

$$1 - S_1 = 1 - \int_{M_{r_1}} p(x) d\mu(x) = \int_{U_{r_1}} p(x) d\mu(x) \le c_2 \mu(U_{r_1}) = c_2 T_1. \tag{129}$$

From these, we can get

$$\begin{aligned}
S_1 + c_1 T_1 &\ge c_1, && (130)\\
S_1 + c_2 T_1 &\ge 1. && (131)
\end{aligned}$$

https://doi.org/10.52202/079017-0329

As $c_1 < c_1 e^\epsilon < c_2$, we can express $c_1 e^\epsilon$ as a convex combination of c_1 and c_2, as

$$c_1 e^\epsilon = \frac{c_2 - c_1 e^\epsilon}{c_2 - c_1} c_1 + \frac{c_1 e^\epsilon - c_1}{c_2 - c_1} c_2. \tag{132}$$

Hence, by taking $\frac{c_2 - c_1 e^\epsilon}{c_2 - c_1}$[equation (130)] + $\frac{c_1 e^\epsilon - c_1}{c_2 - c_1}$[equation (131)], we get

$$\begin{aligned} S + c_1 e^\epsilon T &\geq \frac{c_2 - c_1 e^\epsilon}{c_2 - c_1} c_1 + \frac{c_1 e^\epsilon - c_1}{c_2 - c_1} && (133)\\ &= \frac{c_2 - c_1 e^\epsilon + e^\epsilon - 1}{c_2 - c_1} c_1 && (134)\\ &= \frac{(e^\epsilon - 1)(1 - c_1) + c_2 - c_1}{c_2 - c_1} c_1 && (135)\\ &= r_1. && (136) \end{aligned}$$

Thus we have $\int q_{r_1}(x) d\mu(x) \geq 1$.

Similarly, we show that $\int q_{r_2}(x) d\mu(x) \leq 1$. from $p(x) \leq c_2$, we have $\frac{1}{r_2} p(x) \leq \frac{c_2}{r_2} = b e^\epsilon$. Hence $U_{r_2} = \emptyset$. Thus

$$\begin{aligned} \int q_{r_2}(x) d\mu(x) &= b\mu(L_{r_2}) + \frac{1}{r_2} \int_{M_{r_2}} p(x) d\mu(x) && (137)\\ &= \frac{1}{r_2} \left(c_2 e^{-\epsilon} \mu(L_{r_2}) + \int_{M_{r_2}} p(x) d\mu(x) \right). && (138) \end{aligned}$$

Similar as above, let $S_2 = \mu(L_{r_2})$ and $T_2 = \int_{M_{r_2}} p(x) d\mu(x)$, so that

$$\int q_{r_2}(x) d\mu(x) = \frac{1}{r_2} (c_2 e^{-\epsilon} S_2 + T_2), \tag{139}$$

and, we have

$$T_2 = \int_{M_{r_2}} p(x) d\mu(x) \leq c_2 \mu(M_{r_2}) = c_2 (1 - \mu(L_{r_2})) = c_2 (1 - S_2) \tag{140}$$

and

$$1 - T_2 = 1 - \int_{M_{r_2}} p(x) d\mu(x) = \int_{L_{r_2}} p(x) d\mu(x) \geq c_1 \mu(L_{r_2}) = c_1 S_2. \tag{141}$$

hence

$$\begin{aligned} c_2 S_2 + T_2 &\leq c_2, && (142)\\ c_1 S_2 + T_2 &\leq 1. && (143) \end{aligned}$$

As $c_1 \leq c_2 e^{-\epsilon} \leq c_2$, we have

$$c_2 e^{-\epsilon} = \frac{c_2 - c_2 e^{-\epsilon}}{c_2 - c_1} c_1 + \frac{c_2 e^{-\epsilon} - c_1}{c_2 - c_1} c_2 \tag{144}$$

and by the same reason, we have

$$\begin{aligned} c_2 e^{-\epsilon} S_2 + T_2 &\leq \frac{c_2 - c_2 e^{-\epsilon}}{c_2 - c_1} + \frac{c_2 e^{-\epsilon} - c_1}{c_2 - c_1} c_2 && (145)\\ &= \frac{1 - e^{-\epsilon} + c_2 e^{-\epsilon} - c_1}{c_2 - c_1} c_2 && (146)\\ &= \frac{(e^\epsilon - 1)(1 - c_1) + c_2 - c_1}{c_2 - c_1} c_2 e^{-\epsilon} && (147)\\ &= r_2. && (148) \end{aligned}$$

Thus we have $\int q_{r_2}(x) d\mu(x) \leq 1$.

Finally, we show that $\int q_{r_1}(x)d\mu(x) = 1$ implies $\int q_r(x)d\mu(x) = 1$ for all $r \in [r_1, r_2]$. Let us assume $\int q_{r_1}(x)d\mu(x) = 1$. We first claim that: $p(x) = c_2$ for μ-a.e. $x \in U_{r_1}$, and $p(x) = c_1$ for μ-a.e. $x \in \mathcal{X}\backslash U_{r_1}$. Again, we divide the case of $c_1 = 0$ and $c_1 > 0$.

Suppose first that $c_1 = 0$. By tracking the proof of $\int q_{r_1}(x)d\mu(x) \geq 1$, We can observe that the equality $\int q_{r_1}(x)d\mu(x) = 1$ holds if and only if $\mu(U_0) = 1/c_2$, if and only if $p(x) = c_2$ for μ-a.e. $x \in U_0$. By definition of U_0, we have $p(x) = 0 = c_1$ for all $x \in \mathcal{X}\backslash U_0$. Hence we get the claim.

Next, suppose that $c_1 > 0$. Again, by tracking the proof of $\int q_{r_1}(x)d\mu(x) \geq 1$, We can observe that the equality $\int q_{r_1}(x)d\mu(x) = 1$ is equivalent to any of the following statements:

- Both $S_1 + c_1T_1 = c_1$ and $S_1 + c_2T_1 = 1$ holds.
- Both $\int_{M_{r_1}} p(x)d\mu(x) = c_1\mu(M_{r_1})$ and $\int_{U_{r_1}} p(x)d\mu(x) = c_2\mu(U_{r_1})$ holds.
- $p(x) = c_1$ for μ-a.e. $x \in M_{r_1}$ and $p(x) = c_2$ for μ-a.e. $x \in U_{r_1}$.

Since $L_{r_1} = \emptyset$, we also get the claim.

WLOG, assume that $p(x) = c_2$ for *all* $x \in U_{r_1}$, and $p(x) = c_1$ for *all* $x \in \mathcal{X}\backslash U_{r_1}$. Then, for every $r \in (r_1, r_2]$, we have the following:

- For each $x \in U_{r_1}$, we have $\frac{1}{r}p(x) \geq \frac{c_2}{r_2} = be^{\epsilon}$, hence $q_r(x) = be^{\epsilon}$.
- For each $x \in \mathcal{X}\backslash U_{r_1}$,
  - If $c_1 = 0$, then $p(x) = 0$, $q_r(x) = b$, and
  - If $c_1 > 0$, then $r_1 > 0$, $\frac{1}{r}p(x) < \frac{c_1}{r_1} = b$, hence again $q_r(x) = b$.

Also, we have $1 = \int p(x)d\mu(x) = c_2\mu(U_{r_1}) + c_1(1 - \mu(U_{r_1}))$, hence $\mu(U_{r_1}) = \frac{1-c_1}{c_2-c_1} = \alpha$. It follows that for every $r \in (r_1, r_2]$, we have $\int q_r(x)d\mu(x) = be^{\epsilon}\mu(U_{r_1}) + b(1 - \mu(U_{r_1})) = be^{\epsilon}\alpha + b(1 - \alpha) = 1$. This concludes the proof. □

C.6 Proof of Lemma C.7

Although the proof can be found in [59], we present the proof here for the completeness.

If $r_1 = 0$ and $f(0) = \infty$, then the RHS of the inequality we want to show is ∞, thus it becomes trivial. Hence, we may assume that either $r_1 > 0$ or $f(0) < \infty$.

By the assumption, $p(x)/q(x)$ is the convex combination of r_1 and r_2 for μ-a.e. $x \in \mathcal{X}$, as follows.

$$p(x)/q(x) = \frac{(p(x)/q(x)) - r_1}{r_2 - r_1}r_2 + \frac{r_2 - (p(x)/q(x))}{r_2 - r_1}r_1. \tag{149}$$

Hence, by the convexity of f and $\int p(x)d\mu(x) = \int q(x)d\mu(x) = 1$, we have

$$\begin{aligned} D_f(P\|Q) &= \int q(x)f(p(x)/q(x))d\mu(x) && (150)\\ &\leq \int q(x)\left(\frac{(p(x)/q(x)) - r_1}{r_2 - r_1}f(r_2) + \frac{r_2 - (p(x)/q(x))}{r_2 - r_1}f(r_1)\right)d\mu(x) && (151)\\ &= \int \left(\frac{p(x) - r_1q(x)}{r_2 - r_1}f(r_2) + \frac{r_2q(x) - p(x)}{r_2 - r_1}f(r_1)\right)d\mu(x) && (152)\\ &= \frac{1 - r_1}{r_2 - r_1}f(r_2) + \frac{r_2 - 1}{r_2 - r_1}f(r_1). && (153) \end{aligned}$$

□

C.7 Proof of Lemma C.9

First, we claim that for each $i \in [t - u]$, we can construct a partition $\{B_{i,1}, B_{i,2}, \cdots, B_{i,u}\}$ of A_{u+i} into u (measurable) sets, such that for each $j \in [u]$, the sets $A_j, B_{1,j}, B_{2,j}, \cdots, B_{t-u,j}$ are disjoint. The construction is in the inductive way as follows: Given $i \in [t-u]$, suppose that we have constructed

 https://doi.org/10.52202/079017-0329

such partitions of $A_{u+1}, \cdots, A_{u+i-1}$ such that for each $j \in [u]$, $A_j, B_{1,j}, B_{2,j}, \cdots, B_{i-1,j}$ are disjoint. Let $C_{i,j} = A_j \cup \left(\cup_{k=1}^{i-1} B_{k,j}\right)$. Then for each $x \in A_{u+i}$, at least one of $C_{i,1}, C_{i,2}, \cdots, C_{i,u}$ does not contain x, because

$$\sum_{j=1}^{u} \mathbb{1}_{\mathcal{X}\setminus C_{i,j}}(x) = \sum_{j=1}^{u} \left(1 - \mathbb{1}_{C_{i,j}}(x)\right) = u - \sum_{j=1}^{u} \mathbb{1}_{C_{i,j}}(x) \tag{154}$$

$$= u - \sum_{j=1}^{u} \left(\mathbb{1}_{A_j}(x) + \sum_{k=1}^{i-1} \mathbb{1}_{B_{k,j}}(x) \right) = u - \sum_{j=1}^{u} \mathbb{1}_{A_j}(x) - \sum_{j=1}^{u}\sum_{k=1}^{i-1} \mathbb{1}_{B_{k,j}}(x) \tag{155}$$

$$= u - \sum_{j=1}^{u} \mathbb{1}_{A_j}(x) - \sum_{k=1}^{i-1}\sum_{j=1}^{u} \mathbb{1}_{B_{k,j}}(x) = u - \sum_{j=1}^{u} \mathbb{1}_{A_j}(x) - \sum_{k=1}^{i-1} \mathbb{1}_{A_{u+k}}(x) \tag{156}$$

$$\geq u - (u-1) = 1, \tag{157}$$

where the last line is from that since $x \in A_{u+i}$, at most $u-1$ sets among $A_1, A_2, \cdots, A_{u+i-1}$ can contain x. Hence, setting

$$B_{i,j} = \left\{ x \in A_{u+i} : j = \min\left(\tilde{j} \in [u] : x \notin C_{i,\tilde{j}}\right) \right\} \tag{158}$$

gives the partition $\{B_{i,1}, B_{i,2}, \cdots, B_{i,u}\}$ of A_{u+i}, such that for each $j \in [u]$, $C_{i,j}$ and $B_{i,j}$ are disjoint. This implies that $A_j, B_{1,j}, B_{2,j}, \cdots, B_{i,j}$ are disjoint.

Now, let $\ell = \min_{i\in[t]} Q_i(A_i)$. Using these partitions, we can now show that

$$u = \sum_{j=1}^{u} Q_j(\mathcal{X}) \geq \sum_{j=1}^{u} Q_j\left(A_j \cup \left(\bigcup_{k=1}^{t-u} B_{k,j}\right)\right) \tag{159}$$

$$= \sum_{j=1}^{u} Q_j(A_j) + \sum_{j=1}^{u}\sum_{k=1}^{t-u} Q_j(B_{k,j}) = \sum_{j=1}^{u} Q_j(A_j) + \sum_{k=1}^{t-u}\sum_{j=1}^{u} Q_j(B_{k,j}) \tag{160}$$

$$= \sum_{j=1}^{u} Q_j(A_j) + \sum_{k=1}^{t-u} Q_j(A_{u+k}) \tag{161}$$

$$\geq \sum_{j=1}^{u} Q_j(A_j) + \sum_{k=1}^{t-u} e^{-\epsilon} Q_{u+k}(A_{u+k}) \tag{162}$$

$$\geq \ell\left(u + e^{-\epsilon}(t-u)\right). \tag{163}$$

Hence

$$\ell \leq \frac{u}{u + e^{-\epsilon}(t-u)} = \frac{(u/t)e^{\epsilon}}{(u/t)e^{\epsilon} + 1 - (u/t)}. \tag{164}$$

□

D Proofs of Remaining Propositions

We present the proofs of remaining propositions in the paper, Propositions 3.2 and 4.1.

D.1 Proof of Proposition 3.2

By the assumption, there are subsets $\{A_i\}_{i=1}^{\infty}$ of $\mathcal{X}$ which are pairwise disjoint and $P_i(A_i) = 1$ for all $i \in \mathbb{N}$. Let us pick $P_0 \in \tilde{\mathcal{P}}$, and let $Q_0 = \mathbf{Q}(P_0)$. Since A_i's are disjoint, we have $\sum_{i=1}^{\infty} Q_0(A_i) = Q_0\left(\bigcup_{i=1}^{\infty} A_i\right) \leq 1 < \infty$. Hence, we have $\lim_{i\to\infty} Q_0(A_i) = 0$. Also, by definition of ϵ-LDP, for any $P \in \tilde{\mathcal{P}}$ and $i \in \mathbb{N}$, we have $\mathbf{Q}(P)(A_i) \leq e^{\epsilon} Q_0(A_i)$. Especially, this implies $\mathbf{Q}(P_i)(A_i) \leq e^{\epsilon} Q_0(A_i)$, and thus $\lim_{i\to\infty} \mathbf{Q}(P_i)(A_i) = 0$.

Now, similar to the converse proof in Section C.3, let $\beta_i = \mathbf{Q}(P_i)(A_i)$. Then, the push-forward measures of P_i and $\mathbf{Q}(P_i)$ by the indicator function $\mathbb{1}_A$ are Bernoulli distributions with $\Pr(1) = 1$ and β_i, respectively. By the data processing inequality (Theorem B.2), we have

$$D_f\left(P_i \| \mathbf{Q}(P_i)\right) \geq D_f^{\mathrm{B}}\left(1 \| \beta_i\right). \tag{165}$$

Since $\lim_{i\to\infty} \beta_i = 0$, by continuity of D_f^{B}, we have

$$R_f(\mathbf{Q}) \geq \limsup_{i\to\infty} D_f\left(P_i \| \mathbf{Q}(P_i)\right) \geq \lim_{i\to\infty} D_f^{\mathrm{B}}\left(1 \| \beta_i\right) = D_f^{\mathrm{B}}\left(1 \| 0\right) = M_f, \tag{166}$$

where the last equality is because two Bernoulli distributions with $\Pr(1) = 1$ and $\Pr(1) = 0$, respectively, are mutually singular. Hence, we must have $R_f(\mathbf{Q}) = M_f$. □

D.2 Proof of Proposition 4.1

For generality, we prove that the statement of Proposition 4.1 holds in the general setup in Definition C.1. That is, for the setup in Definition C.1, the mechansim $\mathbf{Q}^* = \mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$ satisfies

$$D_{\mathrm{TV}}\left(\mathbf{Q}^*(P), \mathbf{Q}^*(P')\right) \leq \frac{2}{\max(r_P, r_{P'})} D_{\mathrm{TV}}\left(P, P'\right) \tag{167}$$

for all $P, P' \in \tilde{\mathcal{P}}$. As the mechanisms $\mathbf{Q}^*_{k,\epsilon}$ and $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}$ in the paper are special cases of $\mathbf{Q}^*_{c_1,c_2,\mu,\epsilon}$, a proof in this general setup induces Proposition 4.1.

Now, let us assume the setup in Definition C.1 Let $P, P' \in \tilde{\mathcal{P}}$ be given. WLOG, assume that $r_P \geq r_{P'}$. Let $p = dP/d\mu$, $p' = dP'/d\mu$, and $q(x) = \mathrm{clip}\left(\frac{1}{r_P} p(x); b, be^\epsilon\right)$, $q'(x) = \mathrm{clip}\left(\frac{1}{r_{P'}} p'(x); b, be^\epsilon\right)$, so that $q = d\mathbf{Q}^*(P)/d\mu$ and $q' = d\mathbf{Q}^*(P')/d\mu$. For simplicity, we denote $\mathrm{clip}(x) := \mathrm{clip}(x; b, be^\epsilon)$.

We first note the fact that $\mathrm{clip}(x)$ is monotone increasing and 1-Lipschitz in x. From this and the equivalent expressions of the total variation distance in Appendix B, we have

$$\begin{aligned}
&D_{\mathrm{TV}}\left(\mathbf{Q}^*(P), \mathbf{Q}^*(P')\right) && (168)\\
&= \int_{x:q(x)\geq q'(x)} \left(q(x) - q'(x)\right) d\mu(x) && (169)\\
&= \int_{x:q(x)\geq q'(x)} \left(\mathrm{clip}\left(\frac{1}{r_P} p(x)\right) - \mathrm{clip}\left(\frac{1}{r_{P'}} p'(x)\right)\right) d\mu(x) && (170)\\
&= \int_{x:q(x)\geq q'(x)} \left(\mathrm{clip}\left(\frac{1}{r_P} p(x)\right) - \mathrm{clip}\left(\frac{1}{r_P} p'(x)\right)\right) d\mu(x)\\
&\quad + \int_{x:q(x)\geq q'(x)} \left(\mathrm{clip}\left(\frac{1}{r_P} p'(x)\right) - \mathrm{clip}\left(\frac{1}{r_{P'}} p'(x)\right)\right) d\mu(x) && (171)\\
&\leq \int_{x:q(x)\geq q'(x)} \left|\frac{1}{r_P}\left(p(x) - p'(x)\right)\right| d\mu(x) + 0 && (172)\\
&\leq \frac{1}{r_P} \int_{\mathcal{X}} |p(x) - p'(x)| d\mu(x) = \frac{2}{r_P} D_{\mathrm{TV}}\left(P, P'\right). && (173)
\end{aligned}$$

This ends the proof. □

E Behaviors of Proposed Mechanisms

In this appendix, we present the formal proofs for the behaviors of the proposed mechanisms presented in Sections 3.1 and 3.2.

We first observe that the formula of the optimal worst-case f-divergence in general case

$$\frac{1 - r_1}{r_2 - r_1} f(r_2) + \frac{r_2 - 1}{r_2 - r_1} f(r_1) \tag{174}$$

https://doi.org/10.52202/079017-0329

is the y-coordinate value at $x = 1$ of the line segment joining $(r_1, f(r_1))$ and $(r_2, f(r_2))$. As f is convex, this formula is increasing in r_2 and decreasing in r_1, provided that $r_1 < 1 < r_2$.

Now, let us present the proofs.

- If $f(0) = \infty$ and $\tilde{\mathcal{P}}$ contains two mutually singular distributions, then $\mathcal{R}(\mathcal{X}, \tilde{\mathcal{P}}, \epsilon, f) = \infty$.

  Proof. Let $P_1, P_2 \in \tilde{\mathcal{P}}$ be mutually singular distributions with disjoint supports $A_1, A_2 \subset \mathcal{X}$ respectively (That is, $P_1(A_1) = P_2(A_2) = 1$ and $A_1 \cap A_2 = \emptyset$). Suppose that $\mathbf{Q}$ is an ϵ-LDP sampling mechanism for $(\mathcal{X}, \tilde{\mathcal{P}})$ such that $R_f(\mathbf{Q}) < \infty$. Since $f(0) = \infty$, $D_f(P\|Q) < \infty$, implies $Q \ll P$. Hence, as $D_f(P_i\|\mathbf{Q}(P_i)) < \infty$ and $P_i(A_i^c) = 0$, we have $\mathbf{Q}(A_i^c|P_i) = 0$ for each $i = 1, 2$. As $\mathbf{Q}$ satisfies ϵ-LDP, we have $\mathbf{Q}(A_1^c|P_2) \leq \mathbf{Q}(A_1^c|P_1) = 0$, $\mathbf{Q}(A_1^c|P_2) = 0$. Now, since $A_1 \cap A_2 = \emptyset$, we have $A_1^c \cup A_2^c = \mathcal{X}$. But by the union bound, $1 = \mathbf{Q}(\mathcal{X}|P_2) \leq \mathbf{Q}(A_1^c|P_2) + \mathbf{Q}(A_2^c|P_2) = 0$, which is a contradiction. Hence, $R_f(\mathbf{Q}) = \infty$ for every ϵ-LDP sampling mechanism $\mathbf{Q}$. □

From now, assume $f(0) < \infty$.

Let us first prove the behaviors for the case of finite $\mathcal{X}$ in Section 3.1.

- $\mathcal{R}([k], \mathcal{P}([k]), \epsilon, f)$ is decreasing in ϵ and increasing in k.

  Proof. Recall from Appendix C.4.1 that the case of $\mathcal{X} = [k]$, $\tilde{\mathcal{P}} = \mathcal{P}([k])$ can be fit into the general case with

  $$r_1 = 0, \tag{175}$$

  $$r_2 = \frac{e^\epsilon + k - 1}{e^\epsilon}. \tag{176}$$

  Here, r_2 is decreasing in ϵ and increasing in k. As (174) is increasing in r_2, we get the desired claim. □

- For a fixed k, we have $\mathcal{R}([k], \mathcal{P}([k]), \epsilon, f) \to 0$ as $\epsilon \to \infty$.

  Proof. As $\epsilon \to \infty$, we have $r_2 \to 1$. As f is continuous, $f(1) = 0$, and $f(0) < \infty$, we obtain from (174) that

  $$\mathcal{R}([k], \mathcal{P}([k]), \epsilon, f) \to \frac{1-0}{1-0} f(1) + \frac{1-1}{1-0} f(0) = 0 \tag{177}$$

  as $\epsilon \to \infty$. □

- For a fixed k, as $\epsilon \to 0$, we have $\mathbf{Q}^*_{k,\epsilon}(x|P) \to 1/k$ for every $P \in \mathcal{P}([k])$ and $x \in [k]$.

  Proof. We know that $\frac{1}{e^\epsilon + k - 1} \leq \mathbf{Q}^*_{k,\epsilon}(x|P) \leq \frac{e^\epsilon}{e^\epsilon + k - 1}$. As $\epsilon \to 0$, both of $\frac{1}{e^\epsilon + k - 1}$ and $\frac{e^\epsilon}{e^\epsilon + k - 1}$ converges to $\frac{1}{k}$, hence we get the desired claim. □

Next, let us prove the behaviors for the continuous case in Section 3.2.

- If $c_1 = 0$ and $f(0) = \infty$, then $\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f) = \infty$.

  Proof. Since $r_2 > 1$ and $r_1 = 0$, we have $\frac{r_2 - 1}{r_2 - r_1} = \frac{r_2 - 1}{r_2} > 0$. Hence $\frac{r_2 - 1}{r_2 - r_1} f(r_1) = \frac{r_2 - 1}{r_2 - r_1} f(0) = \infty$, which proves $\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f) = \infty$. □

- For a fixed (c_1, c_2), $\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f)$ is decreasing in ϵ.

  Proof. We can observe that r_1 is increasing in ϵ and r_2 is decreasing in ϵ. Since (174) is increasing in r_2 and decreasing in r_1, we get the desired claim. □

- For a fixed (c_1, c_2) with $c_1 = 0$, as $\epsilon \to \infty$, we have $\mathcal{R}(\mathbb{R}^n, \tilde{\mathcal{P}}_{c_1,c_2,h}, \epsilon, f) \to 0$.

  Proof. We can observe that $r_1 = 0$ and $r_2 \to 1$. Hence, by the same argument as (177), we have the desired claim. □

F Detailed Explanation of Setups in Numerical Results

We present details about the setups in producing numerical results in Section 5, and generating Figure 1 in Section 1.

In this appendix, for each $\mu \in \mathbb{R}^n$, $\mathcal{N}_{\mu,\Sigma}[x] = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma)}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu)}{2}\right)$ is the pdf of the n-dimensional Gaussian distribution with mean μ and covariance Σ. Note that we denote $\mathcal{N}(\mu, \Sigma)$ to refer the Gaussian distribution itself with mean μ and covariance matrix Σ.

F.1 Explanation for numerical result for finite data space

In this appendix, we show that among the choices of Q_0 in the baseline for $\mathcal{X} = [k]$, choosing Q_0 to be the uniform distribution minimizes $R_f(\mathbf{Q})$, and present the precise value of $R_f(\mathbf{Q})$ for the baseline with uniform Q_0. From now, let us fix k, ϵ, f, and we denote $\mathbf{Q}_{Q_0}$ to mean the baseline with the reference distribution Q_0. Also, let $\delta_x \in \mathcal{P}([k])$ be the point mass at $x \in [k]$. Recall that

$$\mathcal{M}_{\epsilon,Q_0} = \{Q \in \mathcal{P}(\mathcal{X}) : e^{-\epsilon/2}Q_0(A) \le Q(A) \le e^{\epsilon/2}Q_0(A), \quad \forall A \subset \mathcal{X}\}, \tag{178}$$

and we set the baseline to satisfy that

$$D_f\left(P\|\mathbf{Q}_{Q_0}(P)\right) = \inf_{Q\in\mathcal{M}_{\epsilon,Q_0}} D_f\left(P\|Q\right). \tag{179}$$

First, if $Q_0(x) = 0$ for some $x \in [k]$, then $\mathbf{Q}_{Q_0}(x|P) \le e^{\epsilon/2}Q_0(x) = 0$, $\mathbf{Q}_{Q_0}(x|P) = 0$ for every $P \in \mathcal{P}([k])$. Especially, $\mathbf{Q}_{Q_0}(x|\delta_x) = 0$, and this implies that $\mathbf{Q}_{Q_0}(\delta_x)$ and δ_x are mutually singular. Hence,

$$R_f(\mathbf{Q}_{Q_0}) \ge D_f\left(\delta_x\|\mathbf{Q}_{Q_0}(\delta_x)\right) = M_f, \tag{180}$$

concluding that $R_f(\mathbf{Q}_{Q_0}) = M_f$. Hence, to minimize $R_f(\mathbf{Q}_{Q_0})$, it suffices to set $Q_0(x) > 0$ for all $x \in [k]$. Hence from now, we only consider such Q_0.

We note that for every $x \in [k]$ and $P \in \mathcal{P}([k])$, we have we have $\mathbf{Q}_{Q_0}(x|P) \le e^{\epsilon/2}Q_0(x)$, and furthermore

$$\begin{aligned}
\mathbf{Q}_{Q_0}(x|P) &= 1 - \sum_{y\in[k]\setminus\{x\}} \mathbf{Q}_{Q_0}(y|P) && (181)\\
&\le 1 - \sum_{y\in[k]\setminus\{x\}} e^{-\epsilon/2}Q_0(y) && (182)\\
&= 1 - e^{-\epsilon/2}(1 - Q_0(x)) && (183)\\
&= e^{-\epsilon/2}Q_0(x) + (1 - e^{-\epsilon/2}). && (184)
\end{aligned}$$

Letting

$$B(t) = \min\{e^{\epsilon/2}t, e^{-\epsilon/2}t + (1 - e^{-\epsilon/2})\}, \tag{185}$$

we have

$$\mathbf{Q}_{Q_0}(x|P) \le B(Q_0(x)). \tag{186}$$

Note that $B(t)$ is increasing in t. Now, for any $x \in [k]$, we have

$$\begin{aligned}
R_f(\mathbf{Q}_{Q_0}) &\ge D_f\left(\delta_x\|\mathbf{Q}_{Q_0}(\delta_x)\right) && (187)\\
&= \mathbf{Q}_{Q_0}(x|\delta_x) f\left(\frac{1}{\mathbf{Q}_{Q_0}(x|\delta_x)}\right) + \sum_{y\in[k]\setminus\{x\}} \mathbf{Q}_{Q_0}(y|\delta_x) f(0) && (188)\\
&= \mathbf{Q}_{Q_0}(x|\delta_x) f\left(\frac{1}{\mathbf{Q}_{Q_0}(x|\delta_x)}\right) + (1 - \mathbf{Q}_{Q_0}(x|\delta_x)) f(0). && (189)
\end{aligned}$$

We can observe that the last term can be written in the form of the optimal worst-case f-divergence (174) with $r_1 = 0$ and $r_2 = 1/\mathbf{Q}_{Q_0}(x|\delta_x)$. In other words, let

$$\mathfrak{R}(r_1, r_2) = \frac{1 - r_1}{r_2 - r_1} f(r_2) + \frac{r_2 - 1}{r_2 - r_1} f(r_1). \tag{190}$$

https://doi.org/10.52202/079017-0329

Then we have $R_f(\mathbf{Q}_{Q_0}) \geq \mathfrak{R}(0, 1/\mathbf{Q}_{Q_0}(x|\delta_x))$.

As noted in Appendix E, $\mathfrak{R}(r_1, r_2)$ is increasing in r_2 for $0 \leq r_1 < 1 < r_2$. Since $\mathbf{Q}_{Q_0}(x|\delta_x) \leq B(Q_0(x))$, we have

$$R_f(\mathbf{Q}_{Q_0}) \geq \mathfrak{R}(0, 1/B(Q_0(x))). \tag{191}$$

Since this should hold for all $x \in [k]$, we have

$$R_f(\mathbf{Q}_{Q_0}) \geq \max_{x\in[k]} \mathfrak{R}(0, 1/B(Q_0(x))). \tag{192}$$

Since $\min_{x\in[k]} Q_0(x) \leq 1/k$ for all $Q_0 \in \mathcal{P}([k])$, and $B(t)$ is increasing in t, we have

$$R_f(\mathbf{Q}_{Q_0}) \geq \mathfrak{R}(0, 1/B(1/k)). \tag{193}$$

Now, let μ_k be the uniform distribution over $[k]$. We will show that

$$R_f(\mathbf{Q}_{\mu_k}) = \mathfrak{R}(0, 1/B(1/k)), \tag{194}$$

which suffices to prove that $Q_0 = \mu_k$ minimizes $R_f(\mathbf{Q}_{Q_0})$.

We observe that $\mathcal{M}_{\epsilon,\mu_k}$ is a convex set in $\mathcal{P}([k])$. Since $D_f(P\|Q)$ is jointly convex in (P, Q), we obtain that $D_f(P\|\mathbf{Q}_{\mu_k}(P)) = \min_{Q\in\mathcal{M}_{\epsilon,\mu_k}} D_f(P\|Q)$ is convex in P by [64, Section 3.2.5]. Hence, the maximum of $D_f(P\|\mathbf{Q}_{\mu_k}(P))$ occurs when P is one of the extreme points of $\mathcal{P}([k])$, that is, the point masses δ_x. By the same arguments as in (187)-(189), we have $D_f(\delta_x\|Q) = \mathfrak{R}(0, 1/Q(x))$, and hence

$$\begin{aligned}
R_f(\mathbf{Q}_{\mu_k}) &= \sup_{P\in\mathcal{P}([k])} D_f(P\|\mathbf{Q}_{\mu_k}(P)) && (195)\\
&= \max_{x\in[k]} D_f(\delta_x\|\mathbf{Q}_{\mu_k}(\delta_x)) && (196)\\
&= \max_{x\in[k]} \inf_{Q\in\mathcal{M}_{\epsilon,\mu_k}} D_f(\delta_x\|Q) && (197)\\
&= \max_{x\in[k]} \inf_{Q\in\mathcal{M}_{\epsilon,\mu_k}} \mathfrak{R}(0, 1/Q(x)) && (198)\\
&= \mathfrak{R}\left(0, \frac{1}{\min_{x\in[k]} \sup_{Q\in\mathcal{M}_{\epsilon,\mu_k}} Q(x)}\right). && (199)
\end{aligned}$$

Also, by the same arguments as in (181)-(184), we have

$$Q(x) \leq B(1/k), \quad \forall Q \in \mathcal{M}_{\epsilon,\mu_k}, x \in [k]. \tag{200}$$

Now, we will show that $\sup_{Q\in\mathcal{M}_{\epsilon,\mu_k}} Q(x) = B(1/k)$ for any $x \in [k]$. To show this, we will prove that there is a distribution $Q \in \mathcal{P}([k])$ such that $Q(x) = B(1/k)$ and $Q(y) = \frac{1-B(1/k)}{k-1}$ for all $y \in [k]\backslash\{x\}$, and this Q is contained in $\mathcal{M}_{\epsilon,\mu_k}$. It suffices to show the followings:

$$\frac{1}{k}e^{-\epsilon/2} \leq B(1/k) \leq \min\left\{1, \frac{1}{k}e^{\epsilon/2}\right\}, \tag{201}$$

$$\frac{1}{k}e^{-\epsilon/2} \leq \frac{1 - B(1/k)}{k-1} \leq \min\left\{1, \frac{1}{k}e^{\epsilon/2}\right\}. \tag{202}$$

We note that whenever $0 \leq t \leq 1$, we have

- $B(t) = \min\{e^{\epsilon/2}t, e^{-\epsilon/2}t + (1 - e^{-\epsilon/2})\} \geq t$, and
- $B(t) \leq e^{-\epsilon/2}t + (1 - e^{-\epsilon/2}) = 1 - (1 - t)e^{-\epsilon/2} \leq 1$.

From these, we can easily observe that $B(1/k) \leq \min\left\{1, \frac{1}{k}e^{\epsilon/2}\right\}$, and

$$\frac{1}{k}e^{-\epsilon/2} \leq \frac{1}{k} \leq B(1/k). \tag{203}$$

Also,

$$\frac{1-B(1/k)}{k-1} \leq \frac{1-(1/k)}{k-1} = \frac{1}{k} \leq \min\left\{1, \frac{1}{k}e^{\epsilon/2}\right\}, \tag{204}$$

$$\frac{1-B(1/k)}{k-1} \geq \frac{1-[e^{-\epsilon/2}(1/k)+(1-e^{-\epsilon/2})]}{k-1} = \frac{1}{k}e^{-\epsilon/2}. \tag{205}$$

This shows that $\sup_{Q\in\mathcal{M}_{\epsilon,\mu_k}} Q(x) = B(1/k)$ for any $x \in [k]$. Thus,

$$\min_{x\in[k]} \sup_{Q\in\mathcal{M}_{\epsilon,\mu_k}} Q(x) = B(1/k), \tag{206}$$

$$R_f(\mathbf{Q}_{\mu_k}) = \mathfrak{R}\left(0, \frac{1}{B(1/k)}\right). \tag{207}$$

This ends the proof that $Q_0 = \mu_k$ minimizes $R_f(\mathbf{Q}_{Q_0})$, and $R_f(\mathbf{Q}_{\mu_k}) = \mathfrak{R}(0, 1/B(1/k))$.

F.2 Explanation for the experiment in 1D Gaussian mixture

The precise description of the set of possible client distributions in our experimental setup is as follows:

$$\tilde{\mathcal{P}} = \left\{ P : p(x) = \frac{\sum_{i=1}^k \lambda_i \mathcal{N}_{\mu_i,1}[x]}{\int_{-4}^4 \sum_{i=1}^k \lambda_i \mathcal{N}_{\mu_i,1}[x]dx} \mathbb{1}_{[-4,4]}(x), k \in [K], \lambda_i \geq 0, \sum_{i=1}^k \lambda_i = 1, |\mu_i| \leq 1 \right\}. \tag{208}$$

Each of $P_j \in \tilde{\mathcal{P}}$ is generated by choosing k, λ_i, and μ_i in (208) as follows: (i) First, choose the number of Gaussian distributions k by first sample $\tilde{k}$ from the Poisson distribution with mean k_0 (which is chosen beforehand), and let $k = \min(\tilde{k}+1, K)$, and (ii) after that, sample each of $\mu_1, \cdots, \mu_k$ independently from the uniform distribution on $[-1,1]$, and sample $(\lambda_1, \cdots, \lambda_k)$ from the uniform distribution on $\mathcal{P}([k])$.

The implementation of our proposed mechanism is as follows. We can observe that

$$\int_{-4}^4 \sum_{i=1}^k \lambda_i \mathcal{N}_{\mu_i,1}[x]dx \geq \inf_{\mu\in[-1,1]} \int_{-4}^4 \mathcal{N}_{\mu,1}[x]dx = \int_{-4}^4 \mathcal{N}_{1,1}[x]dx = \Phi(3) - \Phi(-5), \tag{209}$$

where Φ is the cdf of the 1D standard Gaussian distribution, and for each $x \in [-4,4]$, we have

$$\sum_{i=1}^k \lambda_i \mathcal{N}_{\mu_i,1}[x]dx \leq \sup_{\mu\in[-1,1]} \mathcal{N}_{\mu,1}[x] = \frac{1}{\sqrt{2\pi}} \exp\left(-[\max(|x|-1,0)]^2/2\right). \tag{210}$$

Hence, we have $\tilde{\mathcal{P}} \subset \tilde{\mathcal{P}}_{0,1,\tilde{h}} = \tilde{\mathcal{P}}_{0,c_2,h}$, where

$$\tilde{h}(x) = \frac{\exp\left(-[\max(|x|-1,0)]^2/2\right)}{\sqrt{2\pi}(\Phi(3)-\Phi(-5))} \mathbb{1}_{[-4,4]}(x) \tag{211}$$

and $c_2 = \int \tilde{h}(x)dx$, $h(x) = \tilde{h}(x)/c_2$. When implementing our proposed mechanism, we use the bisection method to find a constant r_P. We predetermine the error tolerances $\delta_1, \delta_2 \geq 0$, $\delta_1 < 1$, and we terminate the bisection method to find r_P if the integration of (13) lies in the interval $[1-\delta_1, 1+\delta_2]$. As mentioned in Section 4.1, to consider the error tolerances, we actually implement $\mathbf{Q}^*_{0,c_2,h,\epsilon'}(P)$, with $\epsilon' = \epsilon - \log\frac{1+\delta_2}{1-\delta_1}$.

In the experiment in the paper, we use $k_0 = 2$ and $\delta_1 = \delta_2 = 10^{-5}$.

For the baseline, we use the same hyperparameter setups as in [35, Section 5, Paragraph "Architectures"], except a slight modification in a reference distribution Q_0. In [35], they set the standard Gaussian as the reference distribution, $Q_0 = \mathcal{N}(0,1)$. To consider the truncated domain $[-4,4]$, we set Q_0 as the truncation of the standard Gaussian, that is, Q_0 has a pdf

$$q_0(x) = \frac{\mathcal{N}_{0,1}[x]}{\int_{-4}^4 \mathcal{N}_{0,1}[x]dx} \mathbb{1}_{[-4,4]}(x). \tag{212}$$

https://doi.org/10.52202/079017-0329

The baseline has many other hyperparameters not specified in [35], such as the proposal distribution used for the Metropolis-Hasting algorithm [66, 67], initial model parameters for the weak learner, etc. We noticed that the code for [35] is available at `https://github.com/karokaram/PrivatedBoostedDensities/tree/master` and hence we made every effort to faithfully reproduce the baseline with exactly the same hyperparameters, including those not mentioned in the paper [35]. However, subtle variations may arise due to differences in the programming languages employed; we used Python, while [35] used Julia.

F.3 Explanation for Figure 1

For $k \in \mathbb{N}$ and $\sigma > 0$, the Gaussian ring distribution with k modes and a component-wise variance σ^2 is the mixture of k Gaussian distributions in $\mathbb{R}^2$ with equal weights, where each Gaussian distribution has the covariance $\sigma^2 I_2$ and equally spaced mean in the unit circle, and one of the Gaussian distribution has mean $(1, 0)$. That is, it has a density $\frac{1}{k}\sum_{i=1}^{k} \mathcal{N}_{\mu_i, \sigma^2 I_2}[x]$, with $\mu_i = (\cos\frac{2\pi i}{k}, \sin\frac{2\pi i}{k})$. In Figure 1, the left image is the pdf of the Gaussian ring distribution with $k = 3$ and $\sigma^2 = 0.5$. It is used as a client's original distribution.

For our proposed mechanism, similar to Section 3.2, we use the setup that $\tilde{\mathcal{P}}$ consists of Gaussian mixtures, where each Gaussian has mean within a unit ball centered at the origin and has covariance $0.5 I_2$. that is, $\tilde{\mathcal{P}} = \{\sum_{i=1}^{k} \lambda_i \mathcal{N}(x_i, \sigma^2 I_2) : k \in \mathbb{N}, \lambda_i \geq 0, \sum_{i=1}^{k} \lambda_i = 1, \|x_i\| \leq 1\}$, with $\sigma^2 = 0.5$. We can also observe that $\tilde{\mathcal{P}} \subset \tilde{\mathcal{P}}_{0,1,\tilde{h}}$ for $\tilde{h}(x) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{[\max(0, \|x\|-1)]^2}{2\sigma}\right)$.

Same as Appendix F.2, we have $\tilde{\mathcal{P}}_{0,1,\tilde{h}} = \tilde{\mathcal{P}}_{0,c_2,h}$, where $c_2 = \int \tilde{h}(x)dx$, $h(x) = \tilde{h}(x)/c_2$. Hence, we use $\mathbf{Q}^*_{0,c_2,h,\epsilon}$. The implementation of $\mathbf{Q}^*_{c_1,c_2,h,\epsilon}(P)$ is the same as the description in Appendix F.2 with $\delta = 10^{-5}$. For the baseline, we also use MBDE with the same hyperparameter setup as [35, Section 5, Paragraph "Architectures"], with the (untruncated) 2D standard Gaussian as a reference distribution, $Q_0 = \mathcal{N}(0, I_2)$.

G Additional Numerical Results for Finite Space

In this appendix, we present more numerical results for the finite space case explained in Section 5.1 for other k. We present the result for $k = 5, 20$, and 100 in Figures 5-7. As shown by the figures, the proposed mechanism always has the smaller worst-case f-divergence compared to the baseline, as the optimality is proved for the proposed mechanism. However, the performance gap between two mechanisms decreases as ϵ becomes smaller and k becomes larger.

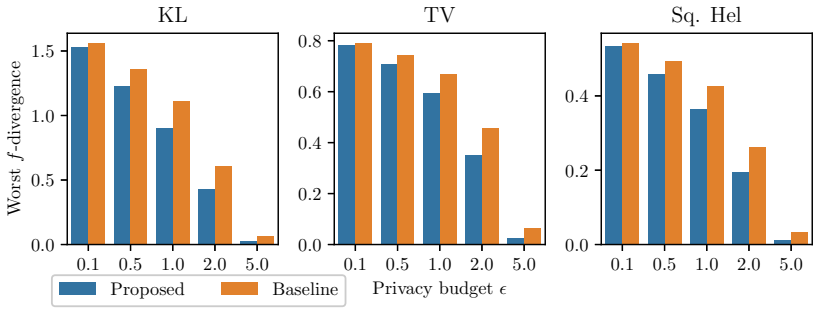

Figure 5: Theoretical worst-case f-divergences of proposed and previously proposed baseline mechanisms (with uniform Q_0) over finite space ($k = 5$)

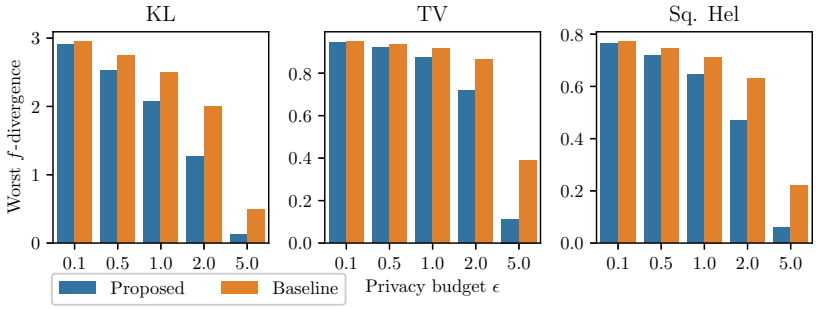

Figure 6: Theoretical worst-case f-divergences of proposed and previously proposed baseline mechanisms (with uniform Q_0) over finite space ($k = 20$)

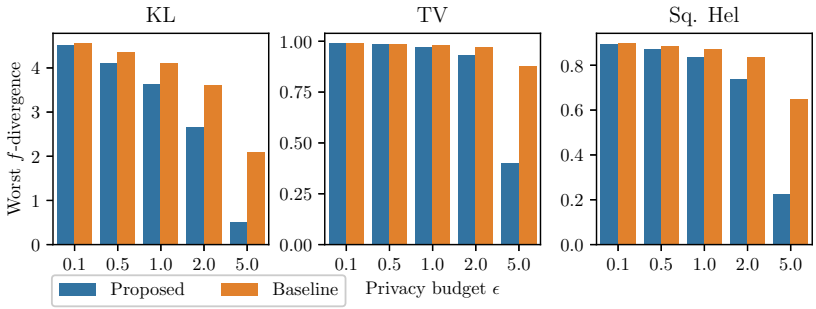

Figure 7: Theoretical worst-case f-divergences of proposed and previously proposed baseline mechanisms (with uniform Q_0) over finite space ($k = 100$)

https://doi.org/10.52202/079017-0329

H Instructions for Reproducing Results

In this appendix, we provide instructions for reproducing the experiments and figures in the paper. For a detailed document about the code, refer to the provided file `README.md`. For the tasks consuming a large time, we also specify the running times of such tasks. We do not specify the running times of the short tasks consuming less than 5 seconds. All experiments are performed on our simulation PC with the following specifications:

- OS: Ubuntu 22.04.1
- CPU: Intel(R) Core(TM) i9-9900X
- Memory: 64GB

There are a few remarks:

- Although we expect that the codes can be run and reproduce our results on sufficiently recent versions of Python libraries, we provide the information about the anaconda environment used in the experiment in `environment.yaml` for the completeness.
- The provided codes contain some lines to make the figures use TEX fonts. Running such lines requires the LATEX to be installed in the experimental environment. We can remove the following lines to disable using TEX:

```
matplotlib.use("pgf")

"pgf.texsystem": "pdflatex",
'font.family': 'serif',
'text.usetex': True,
'pgf.rcfonts': False,
'font.serif' : 'Computer Modern Roman',
```

H.1 Instruction for producing Figure 1

Figure 1 can be obtained by running the code `plot_GaussRing.py`, without any program arguments. We measure the running time for initializing the mechanism and and calculating the sampling distribution for the baseline (MBDE [35]) and our proposed mechanism. The measured running times in our environment are as follows: (unit: second)

- Initializing the mechanism
  - Baseline: 0.80
  - Proposed: 1.22
- Calculating the sampling distribution
  - Baseline: 35.19
  - Proposed: 664.18

H.2 Instruction for producing Figure 2

Figure 2 can be obtained by running the code `visualize_finiteSpace.py`, without any program arguments.

H.3 Instruction for producing results for finite space

The results about the theoretical worst-case f-divergence for finite space, Figures 3,5,6,7, can be obtained by running the code `plot_finite.py` with a program argument `--k` to specify the value of k. For example, in the command line, the aforementioned four figures can be generated by the following commands, respectively:

```
python plot_finite.py --k 10
python plot_finite.py --k 5
```

```
python plot_finite.py --k 20
python plot_finite.py --k 100
```

H.4 Instruction for producing results for 1D Gaussian mixture

The experiment of 1D Gaussian mixture in Section 5.2 consists of the following two codes:

1. `exp_1DGaussMix.py`
   This code performs an experiment on a single ϵ. We can provide two program arguments `--eps` and `--seed` to specify the values of ϵ and the random seed, respectively.
2. `plot_1DGaussMix.py`
   This code generate the plot like Figure 4 from the results of `exp_1DGaussMix.py`

The results in the paper, Figure 4, can be obtained as follows:

1. First, run the following commands:

   ```
   python exp_1DGaussMix.py --eps 0.1 --seed 1
   python exp_1DGaussMix.py --eps 0.5 --seed 2
   python exp_1DGaussMix.py --eps 1.0 --seed 3
   python exp_1DGaussMix.py --eps 2.0 --seed 4
   python exp_1DGaussMix.py --eps 5.0 --seed 5
   ```

   These can be run in any order or in parallel. Check that all of the five result files `data_1DGaussMix_eps{eps}.npy` corresponding to five values of ϵ are generated. In our environment, running all of the five commands in parallel consumes 3h 13m 35s.
2. Then, run the code `plot_1DGaussMix.py` without any program arguments.

I Limitations

Our main contribution lies in proposing a minimax-optimal mechanism, and we present several experimental results based on synthetic data to support the superiority of our mechanism. However, since we have not conducted experiments based on real datasets, the analysis of performance in real-world scenarios is insufficient. Also, our PUT formulation in the minimax sense provides optimal utility in the worst case, which may result in reduced *average* utility when prior information is given. Finally, our implementation of the proposed mechanism requires a large amount of running time due to numerical integration, which makes experiments in multidimensional spaces challenging.

J Broader Impacts

Our proposed mechanism can be utilized for privacy protection in the field of generative models, which has recently received significant attention. A major deterrent to the adoption of privacy protection algorithms in real-world scenarios is the potential performance degradation. Our PUT-optimal mechanism, that minimizes the loss of utility given the privacy budget, can alleviate such concerns.

We should note that an LDP mechanism provides a certain level of privacy protection but cannot guarantee complete privacy protection without completely sacrificing utility. Additionally, clients typically provide multiple data points through various channels, and when these data are combined, it can lead to greater privacy leakage [68, 69].

 https://doi.org/10.52202/079017-0329

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We clearly stated the main contributions and scope of our paper in the abstract and introduction (Section 1).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We stated the limitations of our work in Appendix I.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: All theorems, propositions, and lemmas in the paper include all the necessary assumptions, and are accompanied with either the proofs in the appendices or relevant references.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provided instructions to reproduce all the figures and experimental results in Appendix H.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

https://doi.org/10.52202/079017-0329

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We plan to make the code attached in the supplementary material publicly open after the review period. Also, the instructions to reproduce the results are fully specified in Appendix H.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: We presented enough details to implement our proposed mechanism, and we attached the code for implementing both the proposed mechanism and the baseline in the supplementary material.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: As the main contribution is theoretically characterizing the *worst-case* utility, we focused on extracting the worst-case utility in the experiment.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We provided the information about the computer resources and running times for the experiments in Appendix H.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Our research conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discussed both the positive and negative societal impacts in Appendix J.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

https://doi.org/10.52202/079017-0329

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: We do not recognize any risk for misuse.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: All of the codes only use the basic Python libraries like numpy, scipy, torch, and so on. No other external assets are used.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: As mentioned in Appendix H, we documented the details of our codes in the file `README.md`, which is provided by the NeurIPS Code and Data Submission Guidelines.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

https://doi.org/10.52202/079017-0329

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.