Distribution Simulation Under Local Differential Privacy

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

***Abstract*—We investigate the problem of distribution simulation under local differential privacy: Alice and Bob observe sequences X^n and Y^n respectively, where Y^n is generated by a non-interactive ε-locally differentially private (LDP) mechanism from X^n. The goal is for Alice and Bob to output U and V from a joint distribution that is close in total variation distance to a target distribution P_{UV}. As the main result, we show that such task is impossible if the hypercontractivity coefficient of P_{UV} is strictly bigger than $\left(\frac{e^\varepsilon-1}{e^\varepsilon+1}\right)^2$. The proof of this result also leads to a new operational interpretation of LDP mechanisms: if Y is an output of an ε-LDP mechanism with input X, then the probability of correctly guessing $f(X)$ given Y is bigger than the probability of blind guessing only by $\frac{e^\varepsilon-1}{e^\varepsilon+1}$, for any deterministic finitely-supported function f. If $f(X)$ is continuous, then a similar result holds for the minimum mean-squared error in estimating $f(X)$ given Y.**

I. Introduction

A major challenge in today's machine learning applications is to learn from data as accurately as possible while maintaining the privacy of individuals from whom data is obtained. In such applications, privacy is often quantified in terms of differential privacy [1], that comes in several variants such as approximate DP [2] and Rényi DP [3]. Arguably, the most stringent variant is local DP (LDP), (partially) introduced by Warner [4] in 1960's and formally defined around forty years later in [5, 6]. Informally speaking, a mechanism (or a channel) is locally differentially private if it's output distribution does not vary significantly by changing the inputs. More precisely, a mechanism is said to be ε-locally differentially private (or ε-LDP for short) if the log-likelihood ratio of the output for two different input is smaller than ε almost surely.

Suppose Alice and Bob observe samples $X^n = (X_1, X_2, \dots, X_n)$ and Y^n respectively, where $\{(X_i, Y_i)\}_{i=1}^n$ are drawn i.i.d. from P_{XY}. The goal is for Alice and Bob to apply some (possibly randomized) function to their observation and generate respectively U_n and V_n, whose joint distribution asymptotically approximates a given P_{UV} on $\mathcal{U} \times \mathcal{V}$. This problem, formally referred to as *distribution simulation*, was initiated by Gács and Körner [7] and Wyner [8], and further studied more recently by Kamath and Anantharam [9]. Distribution simulation also generalizes *correlation distillation* [10, 11] in which the goal is to maximize the probability of agreement between Alice and Bob. Despite all progresses, characterizing families of distributions P_{UV} that can be simulated by a given P_{XY} remains challenging, even when P_{XY} and P_{UV} are extremely simple. For instance, for P_{XY} being uniform on $\{(0,0),(0,1),(1,0)\}$ and U and V being uniform random variables on $\{-1,+1\}$ with correlation, say, 0.49, it is still open whether P_{UV} can be simulated using P_{XY}.

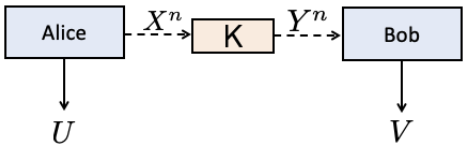

Fig. 1. Alice observes X^n and she applies a locally differentially private mechanism K to her observation to generate Y^n. After transmitting Y^n to a distant party, say, Bob, the goal is for Alice and Bob to generate U and V respectively such that (U, V) is drawn from a target distribution P_{UV}.

In this work, we study the distribution simulation under local differential privacy constraint. That is, we assume each Y_i, $i \in [n] \coloneqq \{1, \dots, n\}$, is the output of an ε-LDP mechanism K with the input X_i. In information-theory parlance, X^n and Y^n are the input and output of the memoryless channel $P_{Y^n|X^n}(y^n|x^n) = \prod_{i=1}^n \mathsf{K}(y_i|x_i)$ and K is assumed to be ε-LDP. This goal is depicted in Fig. 1. In the DP literature, this setting is usually called *non-interactive mechanism* [12].

As the main contribution, we derive a nearly tight upper bound for the hypercontractivity coefficient [13] of input and output of a locally differentially private mechanism. For any joint distribution P_{XY}, the hypercontractivity coefficient $s(X;Y)$ is defined as the supremum of $\frac{D(Q_Y\|P_Y)}{D(Q_X\|P_X)}$, where the supremum is taken over all distributions Q_X on $\mathcal{X}$, not equal to P_X, and Q_Y is the corresponding distribution on $\mathcal{Y}$. This quantity plays an important role in analysis, probability theory, information theory, and discrete Fourier analysis. Interested readers can refer to [10, 11, 14] for a brief summary of their development and impact in these areas. In Theorem 1, we show that if X and Y are input and output of an ε-LDP mechanism, then $s(X;Y) \le \Upsilon_\varepsilon^2$, where $\Upsilon_\varepsilon \coloneqq \frac{e^\varepsilon-1}{e^\varepsilon+1}$. Combining this result with some well-known properties of $s(X;Y)$, we then present an impossibility result for the distribution simulation under local differential privacy in terms of the hypercontractivity coefficient. We demonstrate that if $S(U;V) > \Upsilon_\varepsilon$, then the

simulation of P_{UV} using X^n and Y^n is impossible.

We instantiate our result for an important and insightful joint distribution, namely, U and V are uniform on $\{-1,+1\}$ with correlation $\mathbb{E}[UV]=\rho$. This distribution is called *doubly symmetric binary source* and is denoted by $\mathsf{DSBS}(\rho)$. Notice that $\rho=1$ if and only if $U=V$ and $\rho=0$ if and only if U and V are independent. Thus, ρ can be viewed as a proxy of how likely Alice and Bob can agree on a single bit, that is, $\Pr(U=V)=\frac{1}{2}+\frac{\rho}{2}$. Gács and Körner [7] characterized a necessary and sufficient condition on P_{XY} for simulating $\mathsf{DSBS}(1)$. Applying their result, one can directly show that $\rho<1$ under non-trivial privacy constraint $\varepsilon<\infty$. Thus, a natural question is: What is the maximum correlation $\rho<1$ such that $\mathsf{DSBS}(\rho)$ can be simulated by the input and output of an arbitrary ε-LDP mechanism? We show that the answer to this question is $\frac{e^\varepsilon-1}{e^\varepsilon+1}$ and demonstrate that it is tight.

As a side result, we also present a new operational interpretation of the statistical guarantees provided by local differential privacy. We show that if X and Y are input and output of an ε-LDP mechanism, then reconstructing any deterministic function of X is statistically nearly the same with or without observing Y for reasonably small ε. More specifically, we show that the probability of correctly guessing $f(X)$ for any deterministic finitely-supported function f given Y is upper bounded by $\Upsilon_\varepsilon+\max_a \Pr(f(X)=a)$, where the second term corresponds to the blind guessing. We prove this result by connecting $s(X;Y)$ with the Rényi's maximal correlation [15], which is known to be closely related to the statistical guessing efficiency [16]. If $f(X)$ is continuous, then the same interpretation holds: the minimum mean-squared error (MMSE) in estimating $f(X)$ given Y is slightly smaller than the error corresponding to blind estimation. See Section IV for more details. These result, in addition to the original interpretation of DP in [17], rigorously justify the resulting privacy guarantees, making a better case for adoption of DP technologies in industry.

Notation. For a random variable X, we write P_X and $\mathcal{X}$ for its distribution (i.e., $X\sim P_X$) and its alphabet, respectively. For any set A, we denote by $\mathcal{P}(A)$ the set of all probability distributions on A. Given two sets $\mathcal{X}$ and $\mathcal{Y}$, a mechanism (i.e., channel) K is a mapping from $\mathcal{X}$ to $\mathcal{P}(\mathcal{Y})$ given by $x\mapsto \mathsf{K}(\cdot|x)$. Given $P\in\mathcal{P}(\mathcal{X})$ and a mechanism $\mathsf{K}:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, we let $P\otimes\mathsf{K}$ and $P\mathsf{K}$ denote the corresponding joint distribution $P_{XY}(x,y)=P(x)\mathsf{K}(y|x)$ and output distribution $P\mathsf{K}(\cdot)=\int \mathsf{K}(\cdot|x)P(\mathrm{d}x)$, respectively. We use $\mathsf{Bernoulli}(\alpha)$ for the Bernoulli distribution on $\{-1,+1\}$ with parameter α and $\mathsf{BSC}(\omega)$ for the binary symmetric channel with crossover probability ω. Also, we write $\mathsf{DSBS}(\rho)$ for the joint distribution $\mathsf{Bernoulli}(\frac{1}{2})\otimes\mathsf{BSC}(\frac{1-\rho}{2})$.

II. Preliminaries and Problem Formulation

In this section, we first give a set of definitions that are required for the subsequent sections and also describe the problem formulation.

A. f-Divergences

Given a convex function $f:(0,\infty)\to\mathbb{R}$ such that $f(1)=0$, the f-divergence between two probability measures $P\ll Q$ is defined as [18, 19]

$$D_f(P\|Q)\coloneqq \mathbb{E}_Q\Big[f\big(\frac{\mathrm{d}P}{\mathrm{d}Q}\big)\Big]. \tag{1}$$

Due to convexity of f, we have $D_f(P\|Q)\geq f(1)=0$. If, furthermore, f is strictly convex at 1, then equality holds if and only $P=Q$. Examples of f-divergences needed in this paper includes:

- KL-divergence $D(P\|Q)\coloneqq D_f(P\|Q)$ for $f(t)=t\log t$,
- Total-variation distance $\mathsf{TV}(P,Q)\coloneqq D_f(P\|Q)$ for $\frac{1}{2}|t-1|$,
- Squared Hellinger distance $H^2(P,Q)\coloneqq D_f(P\|Q)$ for $f(t)=(1-\sqrt{t})^2$,
- E_γ-divergence $\mathsf{E}_\gamma(P\|Q)\coloneqq D_f(P\|Q)$ for $f(t)=(t-\gamma)_+$ for some $\gamma\geq 1$, where $(a)_+\coloneqq\max\{a,0\}$.

All f-divergences are known to satisfy the data-processing inequality. Consequently, for any channel $\mathsf{K}:\mathcal{X}\mapsto\mathcal{P}(\mathcal{Y})$, we have $D_f(P\mathsf{K}\|Q\mathsf{K})\leq D_f(P\|Q)$ for any pair of distributions (P,Q).

B. Local Differential Privacy

Definition 1 (Local differential privacy [4, 5])**.** *A mechanism $\mathsf{K}:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ is said to be ε-locally differentially private (or ε-LDP for short), for $\varepsilon\geq 0$, if*

$$\mathsf{K}(A|x)\leq e^\varepsilon \mathsf{K}(A|x'),$$

for any measurable set $A\subseteq\mathcal{Y}$ and any arbitrary pair of inputs $x,x'\in\mathcal{X}$. Let $\mathcal{Q}_\varepsilon$ denote the collection of all ε-LDP mechanisms.

Invoking the definition of E_γ-divergence, one can equivalently characterize LDP mechanisms as follows:

$$\mathsf{K}\in\mathcal{Q}_\varepsilon \iff \sup_{x,x'\in\mathcal{X}} \mathsf{E}_{e^\varepsilon}(\mathsf{K}(\cdot|x)\|\mathsf{K}(\cdot|x'))=0. \tag{2}$$

This equivalent expression has been instrumental in proving several new results in differential privacy literature [20–25]. The smaller the value of ε is, the stronger privacy guarantee is achieved. In particular, any 0-LDP channel would necessarily generate outputs independent of its input.

C. Problem Formulation

Given independent samples X^n from P, one can construct Y^n by applying K an ε-LDP mechanism n times independently, i.e., Y_i is the output of the channel K with X_i as the input for $i\in[n]\coloneqq\{1,\dots,n\}$. We wish to determine how well a target distribution $P_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V})$ can be simulated by two parties, Alice an Bob, observing X^n and Y^n separately. This objective can be formally defined as follows.

Definition 2 (distribution simulation [9])**.** *Given observation $(X^n,Y^n)\overset{\text{iid}}{\sim}P_{XY}$ and a target joint distribution $P_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V})$, we say that the simulation of P_{UV} using P_{XY}*

is possible if there exists a sequence of (possibly randomized) functions $\{\varphi_n\}_{n\in\mathbb{N}}$ *and* $\{\psi_n\}_{n\in\mathbb{N}}$*:*

$$\varphi_n : \mathcal{X}^n \to \mathcal{U}, \qquad \psi_n : \mathcal{Y}^n \to \mathcal{V},$$

such that $U_n := \varphi_n(X^n)$ *and* $V_n := \psi_n(Y^n)$ *asymptotically approximate* U *and* V *in distribution, that is,* $\mathsf{TV}(P_{UV}, P_{U_nV_n}) \to 0$ *as* $n \to \infty$*.*

According to this definition, the task of simulating P_{UV} by Alice and Bob is performed as follows: Alice applies some function to her observations X^n to generate U_n and Bob applies another function to Y^n to generate V_n such that (U_n, V_n) can asymptotically approximate (U, V).

Of particular interest is when P_{UV} is the distribution where U and V are marginally uniform over $\{-1, +1\}$ and U is a ρ-correlated copy of V, i.e., $\mathbb{E}[UV] = \rho$. This joint distribution, often referred to as *doubly symmetric binary source* and denoted by $\mathsf{DSBS}(\rho)$, can be realized by $U \sim \mathsf{Bernoulli}(\frac{1}{2})$ and V being the output of $\mathsf{BSC}(\frac{1-\rho}{2})$. We seek to answer the following question:

Question: What is the maximum ρ such that the simulation of $\mathsf{DSBS}(\rho)$ using $P \otimes \mathsf{K}$ with $P \in \mathcal{P}(\mathcal{X})$ and $\mathsf{K} \in \mathcal{Q}_\varepsilon$ is possible?

III. Main Results

In this section, we develop a machinery in terms of the hypercontractivity coefficient to answer the question posed in the previous section. In order to expound the main result and proof technique we need the following definition.

Definition 3 ([13])**.** *Given* $(X, Y) \sim P \otimes \mathsf{K}$ *with* $P \in \mathcal{P}(\mathcal{X})$ *and* $\mathsf{K} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$*, we define their hypercontractivity coefficient as*

$$s(X;Y) := \sup_{\substack{Q\in\mathcal{P}(\mathcal{X})\\ Q\neq P}} \frac{D(P\mathsf{K}\|Q\mathsf{K})}{D(P\|Q)},$$

where the supremum is taken over all distributions $Q \in \mathcal{P}(\mathcal{X})$ *not equal to* P*.*

Hypercontractivity coefficient exhibits two important properties. First, it is known that it satisfies the data processing inequality, that is [9, Appendix B]

$$s(U;V) \le s(X;Y), \tag{3}$$

for any pair of random variables (U, V) such that we have the Markov chain $U - X - Y - V$. Second, if (X_i, Y_i) for $i \in [n]$ are independent, then [26]

$$s(X^n;Y^n) = \max_{i\in[n]} s(X_i;Y_i). \tag{4}$$

These two properties are essential to obtain a necessary and sufficient condition for the set of all joint distributions P_{UV} that can be simulated by P_{XY}. In fact, it has been shown [9] that P_{UV} can be simulated using P_{XY} if and only if

$$s(U;V) \le s(X;Y). \tag{5}$$

Therefore, an upper bound on $s(X;Y)$ leads to an upper bound on the hypercontractivity coefficient of P_{UV} than can be simulated by P_{XY}. In the following theorem, we present an upper bound for $s(X;Y)$ when X and Y are input and output of an ε-LDP mechanism, respectively.

Theorem 1. *Let* $(X, Y) \sim P\otimes\mathsf{K}$ *with* $P \in \mathcal{P}(\mathcal{X})$ *and* $\mathsf{K} \in \mathcal{Q}_\varepsilon$*. Then, we have*

$$s(X;Y) \le \Upsilon_\varepsilon^2, \tag{6}$$

where

$$\Upsilon_\varepsilon := \frac{e^\varepsilon - 1}{e^\varepsilon + 1}. \tag{7}$$

Proof Sketch. We provide a proof sketch for a slightly stronger result. For any Markov kernel $\mathsf{K} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, define

$$\eta(\mathsf{K}) := \sup_P s(X;Y) \tag{8}$$

$$= \sup_{\substack{P,Q\in\mathcal{P}(\mathcal{X})\\ P\neq Q}} \frac{D(P\mathsf{K}\|Q\mathsf{K})}{D(P\|Q)}, \tag{9}$$

which is sometimes called the *contraction* coefficient of K [27, 28]. With this definition in place, we prove that $\eta(\mathsf{K}) \le \Upsilon_\varepsilon^2$ for any $\mathsf{K} \in \mathcal{Q}_\varepsilon$, thus showing that (6) holds for any $P \in \mathcal{P}(\mathcal{X})$. To this goal, we shall follow the following two steps:

- Following [29], we argue that for any Markov kernel K, we have
$$\eta(\mathsf{K}) \le \sup_{x,x'\in\mathcal{X}} H^2(x,x')\left[1 - \frac{1}{4}H^2(x,x')\right], \tag{10}$$
where $H^2(x,x') := H^2(\mathsf{K}(\cdot|x), \mathsf{K}(\cdot|x'))$ is the squared Hellinger distance between $\mathsf{K}(\cdot|x)$ and $\mathsf{K}(\cdot|x')$.
- We then show that for any ε-LDP mechanism $\mathsf{K} : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and $x, x' \in \mathcal{X}$, we have
$$H^2(x,x') \le 2\frac{(e^{\varepsilon/2}-1)^2(1-e^{-\varepsilon})}{e^\varepsilon - e^{-\varepsilon}}. \tag{11}$$
Note that the squared Hellinger distance takes values in $[0, 2]$ and the mapping $t \mapsto t(1 - \frac{1}{4}t)$ is increasing on $[0, 2]$. Thus, plugging (11) into (10) leads to an upper bound on $\eta(\mathsf{K})$, which is in fact the desired result $\eta(\mathsf{K}) \le \Upsilon_\varepsilon^2$.

The detailed proof will be given in the longer version. ∎

It is worth noting that the upper bound given in this theorem holds for any input distribution P and ε-LDP mechanism K. To assess the tightness of this result, we consider next the well-known randomized response mechanism.

Example 1. Let $P = \mathsf{Bernoulli}(\frac{1}{2})$ and $\mathsf{K} = \mathsf{BSC}(\frac{1}{1+e^\varepsilon})$. It can be verified that this channel, often called randomized-response mechanism [4], is ε-LDP. Fig. 2 presents the gap in the upper bound given in Theorem 1, i.e., $\Upsilon_\varepsilon^2 D(P\|Q) - D(P\mathsf{K}\|Q\mathsf{K})$ for $Q = \mathsf{Bernoulli}(0.6)$. According to this plot, for this simple mechanism, this gap is smaller than 0.03 for any value of $\varepsilon \ge 0$.

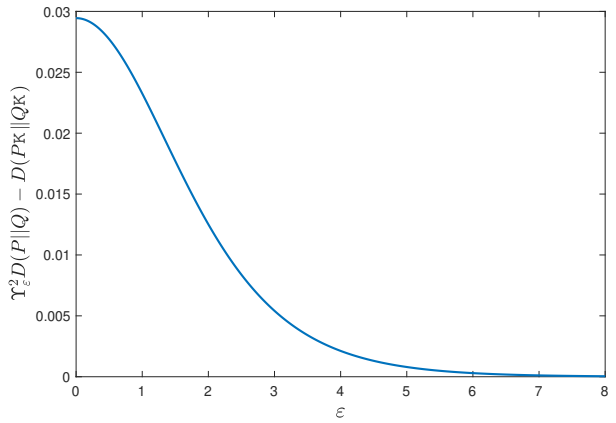

Fig. 2. The gap in the upper bound given in Theorem 1 for $P = \mathsf{Bernoulli}(\frac{1}{2})$, $Q = \mathsf{Bernoulli}(0.6)$, and K being the binary symmetric channel with crossover probability $\frac{1}{1+e^\varepsilon}$.

Combining Theorem 1 with (5), we obtain the following proposition.

Proposition 1. *Let P be a distribution in $\mathcal{P}(\mathcal{X})$ and K be an ε-LDP mechanism. If the simulation of P_{UV} using $P \otimes \mathsf{K}$ is possible, then*

$$s(U,V) \leq \Upsilon_\varepsilon^2,$$

where Υ_ε was defined in (7)*.*

We now instantiate this result for a specific joint distribution P_{UV}.

Corollary 1. *Let P be a distribution in $\mathcal{P}(\mathcal{X})$ and K be an ε-LDP mechanism. If the simulation of $\mathsf{DSBS}(\rho)$ using $P \otimes \mathsf{K}$ is possible, then*

$$\rho \leq \Upsilon_\varepsilon.$$

Proof. It is well-known that the hypercontractivity coefficient $\mathsf{DSBS}(\rho)$ is equal to ρ^2 [9]. The desired result then follows by plugging $s(U;V) = \rho^2$ into Proposition 1. ∎

It is worth mentioning that the upper bound for ρ given in the above corollary is tight. That is, there exist an ε-LDP mechanism K and distribution P in $\mathcal{P}(\mathcal{X})$ such that the simulation of $\mathsf{DSBS}(\rho)$ using $P \otimes \mathsf{K}$ is possible *if and only if* $\rho \leq \Upsilon_\varepsilon$. To observe this, let $P_{XY} = \mathsf{DSBS}(\Upsilon_\varepsilon)$, that can be equivalently expressed as $P \otimes \mathsf{K}$ where $P = \mathsf{Bernoulli}(\frac{1}{2})$ and $\mathsf{K} = \mathsf{BSC}(\frac{1}{1+e^\varepsilon})$. Note that, as already mentioned in Example 1, K is an ε-LDP mechanism. The celebrated Witsenhausen's result in [26] indicates that the simulation of $\mathsf{DSBS}(\rho)$ using $\mathsf{DSBS}(\Upsilon_\varepsilon)$ is impossible when $\rho > \Upsilon_\varepsilon$. Moreover, the simulation is possible if $\rho \leq \Upsilon_\varepsilon$: Alice releases X_1 the first bit of her observation and Bob releases a suitably noisy copy of Y_1. Thus, the simulation of $\mathsf{DSBS}(\rho)$ using $\mathsf{DSBS}(\Upsilon_\varepsilon)$ is possible if and only if $\rho \leq \Upsilon_\varepsilon$. This in turn implies that the maximum ρ such that the simulation of $\mathsf{DSBS}(\rho)$ using $P \otimes \mathsf{K}$ for any $P \in \mathcal{P}(\mathcal{X})$ and $\mathsf{K} \in \mathcal{Q}_\varepsilon$ is possible is indeed equal to Υ_ε.

We end this section by a remark that Proposition 1 can give rise to impossibility results for the private simulation of a wide range of joint distributions. In particular, if the hypercontractivity coefficient of P_{UV} can be either computed in closed-form or lower-bounded, then Proposition 1 can be used to characterize an impossibility result. The hypercontractivity coefficient can be expressed in closed-form, for instance, for binary symmetric channel with non-uniform input, Z-channel, binary erasure channel with uniform input. Moreover, it is known to be lower bounded for general joint distributions by the Rényi maximal correlation [30], which is rather straightforward to compute or approximate for discrete distributions.[1]

IV. An Operational Interpretation for Local Differential Privacy

In this section, we exploit Theorem 1 to give a new interpretation of the privacy guarantee provided by the local differential privacy constraint. To expound our result, we need the following definition.

Definition 4. *Given a pair of random variables (A,B), the advantage of reconstructing A given B is given by*

$$\mathsf{Adv}(A|B) := \max_{g:\mathcal{B}\to\mathcal{A}} \Pr(A = g(B)) - \max_{a\in\mathcal{A}} P_A(a),$$

if A is a discrete random variable and

$$\mathsf{Adv}(A|B) := 1 - \frac{\min_{g:\mathcal{B}\to\mathcal{A}} \mathbb{E}[|A-g(B)|^2]}{\mathsf{var}(A)},$$

if A is an $\mathbb{R}$-valued continuous random variable, where $\mathsf{var}(A)$ denotes the variance of A.

We remark that $\max_{g:\mathcal{B}\to\mathcal{A}} \Pr(A = g(B))$ is usually called the probability of correctly guessing A given B and $\min_{g:\mathcal{B}\to\mathcal{A}} \mathbb{E}[|A-g(B)|^2]$ is often referred to as the minimum mean-squared error (MMSE) in estimating A given observation B. It is instructive to note that $\mathsf{Adv}(A|B)$ in fact quantifies the advantage of the observation B in reconstructing A (i.e., guessing or estimating A depending on the alphabet $\mathcal{A}$). In other words, $\mathsf{Adv}(A|B) = 0$ occurs if B does not contribute to reconstructing A at all, that is, A and B are independent. Similarly, if $\mathsf{Adv}(A|B)$ is small, then it is nearly the same to reconstruct A with or without B.

Notice that any ε-LDP mechanism with $\varepsilon = 0$ generates Y that is statistically independent of its input X, and thus $\mathsf{Adv}(f(X)|Y) = 0$ for any deterministic function f. For any reasonably small $\varepsilon > 0$, it is well expected that $\mathsf{Adv}(f(X)|Y)$ be small as well. However, there has not been any precise estimate of how small it is. Next lemma addresses this issue by presenting an upper bound on $\mathsf{Adv}(f(X)|Y)$ in terms of ε.

Lemma 1. *Let X and Y be the input and output of an ε-LDP mechanism K, respectively. Then, we have*

$$\mathsf{Adv}(f(X)|Y) \leq \Upsilon_\varepsilon,$$

[1] A simple proof of the fact that Rényi maximal correlation lower bounds hypercontractivity coefficient can be found in [14].

for any deterministic function f, where Υ_ε was defined in (7).

In light of this lemma, $\mathsf{Adv}(f(X)|Y)$ is at most linear in ε for sufficiently small $\varepsilon > 0$ *for any deterministic function* f.

Proof sketch. First, assume that X is discrete. It was shown in [16, Theorem 9] that $\mathsf{Adv}(f(X)|Y) \leq \rho_m(X;Y)$, where $\rho_m(X;Y)$ is the Rényi's maximal correlation between X and Y, defined as

$$\rho_m(X;Y) \coloneqq \sup_{f,g} \rho(f(X), g(Y)),$$

where the supremum is taken over all measurable real-valued functions f and g on $\mathcal{X}$ and $\mathcal{Y}$, respectively, and $\rho(\cdot,\cdot)$ denotes the Pearson correlation. It is known that $\rho_m^2(X;Y) \leq s(X;Y)$ [14, Theorem 4], and thus $\rho_m(X;Y) \leq \Upsilon_\varepsilon$ if X and Y are input and output of an ε-LDP mechanism, respectively. For continuous X, we invoke [31, Theorem 1] to show MMSE in estimating $f(X)$ given Y is lower bounded by $\mathsf{var}(f(X))(1-\rho_m^2(X;Y))$, from which the result follows. ∎

REFERENCES

[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory of Cryptography (TCC)*, Berlin, Heidelberg, 2006, pp. 265–284.

[2] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *EUROCRYPT*, S. Vaudenay, Ed., 2006, pp. 486–503.

[3] I. Mironov, "Rényi differential privacy," in *Proc. Computer Security Found. (CSF)*, 2017, pp. 263–275.

[4] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[5] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proc. ACM symp. Principles of Database Systems (PODS)*. ACM, 2003, pp. 211–222.

[6] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM J. Comput.*, vol. 40, no. 3, pp. 793–826, Jun. 2011.

[7] P. Gács and J. Körner, "Common information is far less than mutual information," *Problems of Control and Information Theory*, vol. 2, no. 02, pp. 119–162, 1972.

[8] A. Wyner, "The common information of two dependent random variables," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, 1975.

[9] S. Kamath and V. Anantharam, "On non-interactive simulation of joint distributions," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3419–3435, 2016.

[10] E. Mossel and R. O'Donnell, "Coin flipping from a cosmic source: On error correction of truly random bits," *Random Structures & Algorithms*, vol. 26, no. 4, pp. 418–436, 2005.

[11] E. Mossel, R. O'Donnell, O. Regev, J. E. Steif, and B. Sudakov, "Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality," *Israel Journal of Mathematics,*, vol. 154, no. 1, pp. 299–336, 2006.

[12] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. of IEEE Foundations of Computer Science (FOCS)*, 2013.

[13] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, 12 1976.

[14] V. Anantharam, A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and a data processing inequality," in *2014 IEEE Int. Symp. Inf. Theory*, 2014, pp. 3022–3026.

[15] A. Rényi, "On measures of dependence," *Acta Mathematica Academiae Scientiarum Hungaricae*, vol. 10, pp. 441–451, 1959.

[16] F. P. Calmon, A. Makhdoumi, M. Médard, M. Varia, M. Christiansen, and K. R. Duffy, "Principal inertia components and applications," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5011–5038, 2017.

[17] L. Wasserman and S. Zhou, "A statistical framework for differential privacy," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 375–389, 2010.

[18] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observations," *Studia Sci. Math. Hungar.*, vol. 2, pp. 299–318, 1967.

[19] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of Royal Statistics*, vol. 28, pp. 131–142, 1966.

[20] B. Balle and Y.-X. Wang, "Improving the Gaussian mechanism for differential privacy: Analytical calibration and optimal denoising," in *ICML*, vol. 80, 10–15 July 2018, pp. 394–403.

[21] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," in *NeurIPS*, 2018, pp. 6280–6290.

[22] B. Balle, G. Barthe, M. Gaboardi, and J. Geumlek, "Privacy amplification by mixing and diffusion mechanisms," in *NeurIPS*, 2019, pp. 13 277–13 287.

[23] S. Asoodeh, M. Diaz, and F. P. Calmon, "Privacy analysis of online learning algorithms via contraction coefficients," *arXiv 2012.11035*, 2020.

[24] S. Asoodeh, J. Liao, F. P. Calmon, O. Kosut, and L. Sankar, "Three variants of differential privacy: Lossless conversion and applications," *To appear in Journal on Selected Areas in Information Theory (JSAIT)*, 2021.

[25] S. Asoodeh, M. Diaz, and F. P. Calmon, "Privacy amplification of iterative algorithms via contraction coefficients," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2020.

[26] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM Journal on Applied Mathematics*, vol. 28, no. 1, pp. 100–113, 1975.

[27] Y. Polyanskiy and Y. Wu, "Dissipation of information in channels with input constraints," *IEEE Trans. Inf. Theory*, vol. 62, no. 1, pp. 35–55, Jan 2016.

[28] M. Raginsky, "Strong data processing inequalities andϕ-sobolev inequalities for discrete channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3355–3389, June 2016.

[29] O. Ordentlich and Y. Polyanskiy, "Strong data processing constant is achieved by binary inputs," *IEEE Transactions on Information Theory*, pp. 1–1, 2021.

[30] A. Rényi, "On measures of dependence," *Acta Mathematica Academiae Scientiarum Hungarica*, vol. 10, pp. 441–451, 1959.

[31] S. Asoodeh, F. Alajaji, and T. Linder, "Privacy-aware MMSE estimation," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1989–1993.