Local Differential Privacy and Strong Data Processing Inequalities

Adway Patra
Dept. of Electrical and Computer Engineering
University of Maryland, College Park
apatra@umd.edu

May 20, 2021

Abstract

The problem of local differential privacy involves the design and analysis of privatization kernels that satisfy certain probabilistic guarantees while maintaining a desired level of estimation efficacy. Several works have successfully explored the relationship of a privatizing kernel to the contraction of divergence of probability distributions passed through the kernel. Very recently, an equivalence of an (ϵ, δ)-LDP channel to the contraction of the E_γ divergence has been established. The purpose of this work is to study and understand this equivalence and to explore the possibilities of applying them to newer problems of privacy. In this regard, a new concept of non-homogeneous privacy has been proposed and certain bounds on the estimator performance have been established.

1 Introduction

In this age of big data and machine learning, data has really become the new currency. In every aspect of our everyday lives, every gadget that we interact with is most likely collecting some form of data about us. Naturally, people have started asking questions about how much do the gadget owners know about us, or more importantly, how comfortable are we with them knowing. This leads us to the natural question of data privacy. There are several directions through which this assurance of data privacy can be given to the user by an organization. One of the cryptographic approaches, which has garnered much traction in recent times, is Differential Privacy (DP). Loosely speaking, DP ensures that any query made by an accesor of the data, collected from several individuals, does not reveal much information about the data of any single individual. Although DP somewhat puts to rest the fundamental allegations against individual information leakage via queries, being a centralized model, it still lacks the security against the database holder itself. The notion of Local Differential Privacy (LDP), which is a stricter and decentralized version of DP, provides us with such security guarantees. The basic idea of LDP is to introduce noise at the time of data collection itself instead of at the time of answering queries and hence to protect the privacy even against a malicious collector. The drawback is that LDP typically requires more noise than regular DP and hence the probabilistic guarantees for the correctness of the queries become looser.

The author would like to thank Prof. Alexander Barg, Prof. Prakash Narayan and Sagnik Bhattacharya for illuminating discussions and suggestions.

The noise introducing mechanism of LDP to each individual sample can be modeled as a channel operating in a finite sample space where each sample is chosen i.i.d. from some underlying distribution. The LDP guarantees then translate to certain uniformity requirements on the Markov transition kernel of the channel. From the theory of Markov chains, we know that if $U \to X \to Y$ is a Markov chain, then the random variables satisfy the Data Processing inequality, namely $I(U;Y) \leq I(U;X)$ where $I(\cdot;\cdot)$ denotes the mutual information. From another perspective, for two underlying sample distributions P and Q, the push forward through a stochastic kernel makes the two output distributions more "uniform" and hence decreases their distance measured in terms of some divergence. Such inequalities can often be strengthened to get the class of strong data processing inequalities (SDPIs) and the strength of these inequalities are measured in terms of the contraction coefficients.

These insights lead us to the question: is there a way to describe the operation of an LDP channel on the data in the language of SDPIs? The answer to this question has been studied in a flurry of papers and has recently been made concrete.

The structure of this article is as follows. In Section 2, we recall some fundamental required results on privacy and SDPIs. In Section 3, we explore the relationship between LDP and SDPIs which have been used to analyze the class of problems on private estimation, discussed in Section 4. In Section 5, we introduce the idea of Non-Homogeneous Local Differential Privacy. We present a candidate scheme along with it's performance analysis using the tools of Section 3 and 4. We conclude with some future possibilities and directions in Section 6.

2 Preliminaries

2.1 DP and LDP

A database $\mathcal{D}$ is a collection of finite number of samples from an underlying sample space $\mathcal{X}$. A database $\mathcal{D}'$ is called a neighboring database if $d_H(\mathcal{D}, \mathcal{D}') = 1$.

Definition 1. *A randomized mechanism $\mathcal{K} : \mathcal{X}^n \to \mathcal{Y}$, operating on databases of size n, is called (ϵ, δ)-DP,for $\epsilon \geq 0, \delta \in [0,1]$, if for all neighboring databases $\mathcal{D}, \mathcal{D}'$, we have*

$$Pr(\mathcal{K}(\mathcal{D}) \in S) \leq e^{\epsilon} Pr(\mathcal{K}(\mathcal{D}') \in S) + \delta$$

for all $S \in \sigma(\mathcal{Y})$.

As mentioned before, the DP mechanism $\mathcal{K}$ can be thought of doing the privatization while answering the query so as to confuse an adversary as to the underlying database being $\mathcal{D}$ or $\mathcal{D}'$. This is a centralized model because the privatization mechanism is working on the whole database instead of individual samples. The LDP definition is, in a way, the one-shot $(n = 1)$ version of this.

Definition 2. *A randomized mechanism $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ is called (ϵ, δ) LDP, for $\epsilon \geq 0, \delta \in [0,1]$, if*

$$\sup_{x,x' \in \mathcal{X}} \sup_{S \in \sigma(\mathcal{Y})} [Pr(\mathcal{K}(x) \in S) - e^{\epsilon} Pr(\mathcal{K}(x') \in S)] \leq \delta$$

This decentralization can be done at the user end itself at the time of data collection. For n users each holding a random i.i.d. datapoint X_i, the privatization mechanism is said to be non-interactive if each privatizing kernel $\mathcal{K}_i$ only depends on X_i. It is called interactive if $\mathcal{K}_i$ depends in X_i as well as Y_1^{i-1} where $Y_i = \mathcal{K}(X_i)$.

2.2 SDPIs and Contraction Coefficients

We begin with the following definitions.

Definition 3. *For two distributions on $\mathcal{X}$, P and Q, and a convex function $f : (0, \infty) \to \mathbb{R}^+$ satisfying $f(1) = 0$, the f-divergence of P and Q is defined as*

$$D_f(P||Q) = \mathbb{E}_Q[f(\frac{dP}{dQ})]$$

where $P << Q$, i.e., P is absolutely continuous with respect to Q.

The f-divergence incorporates all other popularly used divergence definitions. For example, taking $f(x) := x \log x$ gives us the KL divergence

$$D_{KL}(P||Q) = \int \log \frac{dP}{dQ} dP$$

, taking $f(x) := 0.5|x - 1|$ gives the total variation distance

$$D_{TV}(P||Q) = \frac{1}{2} \int |dP - dQ| = \sup_{S \subseteq \sigma(\mathcal{X})} |P(S) - Q(S)|$$

and $f(x) := (x - 1)^2$ gives the χ^2-divergence

$$D_{\chi^2}(P||Q) = \int (\frac{dP}{dQ})^2 dQ - 1$$

Another popular divergence which has recently come to limelight in [2] for analysis of LDP is for $f_\gamma(x) := \max\{x - \gamma, 0\}$, known as the E_γ-divergence or the "Hockey-Stick" divergence and given formally by

$$E_\gamma(P||Q) = \frac{1}{2} \int |dP - \gamma dQ| - \frac{1}{2}|\gamma - 1|$$

For a transition probability kernel $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ and a probability distribution P on $\mathcal{X}$, let $P\mathcal{K}(y) = \int \mathcal{K}(y|x)P(dx)$. It is well known that all the above divergences satisfy data processing type inequalities, namely

$$D_f(P\mathcal{K}||Q\mathcal{K}) \leq D_f(P||Q)$$

In fact, in many cases, these inequalities can be further strengthened resulting in the set of strong data processing inequalities

$$D_f(P\mathcal{K}||Q\mathcal{K}) \leq \eta_f(\mathcal{K}) D_f(P||Q)$$

where the contraction coefficient $\eta_f(\mathcal{K})$ is defined as

$$\eta_f(\mathcal{K}) = \sup_{P,Q: 0 < D_f(P||Q) < \infty} \frac{D_f(P\mathcal{K}||Q\mathcal{K})}{D_f(P||Q)}$$

For the popular divergences, the coefficients $\eta_{TV}(\mathcal{K}), \eta_{KL}(\mathcal{K}), \eta_{\chi^2}(\mathcal{K})$ and $\eta_\gamma(\mathcal{K})$ follow similar definitions. The inter-relationships between these contraction coefficients have been well studied in the literature. It is known that

$$\eta_{TV}(\mathcal{K}) = \sup_{x,x'} D_{TV}(\mathcal{K}(\cdot|x), \mathcal{K}(\cdot|x')) \tag{1}$$

which gives a very simple two point characterization of the supremum. Another very useful result is

$$\eta_f(\mathcal{K}) \leq \eta_{TV}(\mathcal{K})$$

, i.e., all f-divergences contract at least as much as the TV distance [4]. Further, on finite alphabets, for any twice differentiable f with $f''(1) > 0$

$$\eta_{\chi^2}(\mathcal{K}) \leq \eta_f(\mathcal{K})$$

[1]. With these definitions we are ready to make a connection with LDP.

3 Relationship to LDP

Let $\mathfrak{K}_{\epsilon,\delta}$ be the set of all (ϵ, δ)-LDP mechanisms with input $\mathcal{X}$ and output $\mathcal{Y}$. One of the first among the contraction results for kernels in $\mathfrak{K}_{\epsilon,\delta}$ were discussed in [6]. For any $(\epsilon, 0)$ LDP mechanism $\mathcal{K}$, the following holds

$$\sup_{x,x' \in \mathcal{X}} D_{KL}(\mathcal{K}(\cdot|x), \mathcal{K}(\cdot|x')) \leq \epsilon(e^\epsilon - 1)$$

which implies

$$D_{KL}(\mathcal{K}P||\mathcal{K}Q) \leq \epsilon(e^\epsilon - 1)$$

by convexity. This was later strengthened by [5]

$$D_{KL}(\mathcal{K}P||\mathcal{K}Q) + D_{KL}(\mathcal{K}Q||\mathcal{K}P) \leq \min\{4, e^{2\epsilon}\}(e^\epsilon - 1)^2 D_{TV}(P||Q)^2$$

For the E_γ divergence, a two point characterization, similar to 1, was found in [3]

$$\eta_\gamma(\mathcal{K}) = \sup_{x,x'} D_\gamma(\mathcal{K}(\cdot|x), \mathcal{K}(\cdot|x'))$$

This generalizes 1 because $E_1(P||Q) = D_{TV}(P||Q)$. Using this the authors in [2] were able to put forward the following equivalence between (ϵ, δ)-LDP and the E_γ contraction coefficient

Theorem 1. *([2], Theorem 1) A kernel $\mathcal{K}$ is (ϵ, δ)-LDP if and only if $\eta_{e^\epsilon}(\mathcal{K}) \leq \delta$ or equivalently*

$$\mathcal{K} \in \mathfrak{K}_{\epsilon,\delta} \quad \iff \quad E_{e^\epsilon}(P\mathcal{K}||Q\mathcal{K}) \leq \delta E_{e^\epsilon}(P||Q) \tag{2}$$

The authors further proved a much more generalized contraction applicable to any f-divergence.

Lemma 2. *([2], Lemma 1) For any $\mathcal{K} \in \mathfrak{K}_{\epsilon,\delta}$ and $\phi(\epsilon, \delta) = 1 - e^{-\epsilon}(1 - \delta)$, $\eta_f(\mathcal{K}) \leq \phi(\epsilon, \delta)$ or equivalently*

$$D_f(P\mathcal{K}||Q\mathcal{K}) \leq \phi(\epsilon, \delta) D_f(P||Q) \tag{3}$$

It was also shown that if one considers the n-fold product distribution, i.e., n i.i.d. samples are generated from the underlying distribution and then privatized through the same (ϵ, δ)-LDP mechanism $\mathcal{K}$ independently then the following holds

$$D_f(P^{\otimes n}\mathcal{K}^{\otimes n}||Q^{\otimes n}\mathcal{K}^{\otimes n}) \leq \phi_n(\epsilon, \delta) D_f(P^{\otimes n}||Q^{\otimes n}) \tag{4}$$

where $\phi_n(\epsilon, \delta) = 1 - e^{-\epsilon n}(1 - \delta)^n$.

4 Applications to Private Estimation

Let $\mathcal{P}(\mathcal{X})$ denote a class of distributions on the sample space $\mathcal{X}$, and let $\theta : \mathcal{P}(\mathcal{X}) \to \Theta$ denote a functional defined on $\mathcal{P}(\mathcal{X})$(We shall use $\mathcal{P}$ whenever $\mathcal{X}$ is clear from context). The space Θ in which the parameter $\theta(P)$ takes values depends on the underlying statistical model. For example, in the case of one-dimensional mean estimation problem, it is a subset of the real line, or for the distribution estimation problem, it is a probability simplex over $\mathbb{R}^{|\mathcal{X}|}$. Let ρ denote a semi-metric on the space $\rho : \Theta \times \Theta \to \mathbb{R}_+$. In the non-private setting, the statistician is given direct access to i.i.d. observations $\{X_i\}_{i=1}^n$ drawn according to some distribution $P \in \mathcal{P}$. The goal is to fine and estimator of $\theta(P)$. We define an estimator $\hat{\theta}$ as a measurable function $\hat{\theta} : \mathcal{X}^n \to \Theta$, whose quality is assessed in terms of the risk

$$\mathbb{E}_P[\rho(\hat{\theta}(X_1^n), \theta(P))]$$

and the target is to minimize it over all $P \in \mathcal{P}$, i.e., the minimax risk is

$$\mathfrak{M}_n(\mathcal{P}, \rho) = \inf_{\hat{\theta}} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\rho(\hat{\theta}(X_1^n), \theta(P))]$$

In the private setting, the statistician is given privatized samples $\{Y_i\}_{i=1}^n$ where Z_i is the privatized version of X_i through $\mathcal{K}_i$. The composite mechanism $\{\mathcal{K}_i\}_{i=1}^n$ can be interactive or non-interactive but each $\mathcal{K}_i$ satisfies (ϵ, δ)-LDP constraint. The target now is to design an estimator $\psi : \mathcal{Y}^n \to \Theta$ and the minimax risk is

$$\mathfrak{M}_n(\mathcal{P}, \rho, \epsilon, \delta) = \inf_{\mathcal{K}_i \in \mathfrak{K}_{\epsilon,\delta} \forall i} \inf_{\psi} \sup_{P \in \mathcal{P}} \mathbb{E}_P[\rho(\psi(Y_1^n), \theta(P))]$$

4.1 Locally Private Le Cam's Method

Le Cam's method is applicable when our set of possible underlying distributions is limited to two: P_0 and P_1. Let V denote that random index of the distribution. Then we have the basic inequality in the non-private setting,

$$P_{err}(V|X_1^n) \geq \frac{1}{2}[1 - D_{TV}(P_0^{\otimes n}||P_1^{\otimes n})]$$

and

$$\mathfrak{M}_n(\mathcal{P}, \rho) \geq \frac{\tau}{2}[1 - D_{TV}(P_0^{\otimes n}||P_1^{\otimes n})] \geq \frac{\tau}{2}[1 - \frac{1}{\sqrt{2}}\sqrt{nD_{KL}(P_0||P_1)}]$$

where τ is such that $\rho(\theta(P_0), \theta(P_1)) \geq 2\tau$.

In the private setting, X_1^n is replaced by Y_1^n and $P_i^{\otimes n}$ gets replaced by the output probability distribution of the composite channel $\mathcal{K}^n = \{\mathcal{K}_i\}_{i=1}^n$. It was shown in [2] that in this case the following lower bound on the minmax error holds.

$$\mathfrak{M}_n(\mathcal{P}, \rho, \epsilon, \delta) \geq \frac{\tau}{2}[1 - \frac{1}{\sqrt{2}}\sqrt{n\phi(\epsilon, \delta)D_{KL}(P_0||P_1)}]$$

Hence, heuristically speaking, the cost of privatization is that it reduces the effective sample size from n to $n(1 - e^{-\epsilon}(1 - \delta))$.

4.2 Locally Private Fano's Method

When the set of distributions is not limited to two but is still finite, we can get a lower bound on the error by using Fano's inequality. Like before, let V denote that random variable denoting the index of the distribution which takes value in $\mathcal{V} := \{1, 2, \cdots, |\mathcal{V}|\}$. We have from the standard Fano's inequality in the non-private setting

$$\mathfrak{M}_n(\mathcal{P}, \rho) \geq \tau \left[1 - \frac{I(V; X_1^n) + \log 2}{\log |\mathcal{V}|}\right]$$

where τ is such that $\rho(\theta(P), \theta(P')) \geq 2\tau$ for every $P, P' \in \mathcal{P}$. For the private setting $I(V; X_1^n)$ gets replaced by $I(V; Y_1^n)$. It was shown in [2] that the following upper bound holds when the privatization mechanism is non-interactive and identical for all samples

$$I(V; Y_1^n) \leq \phi_n(\epsilon, \delta) I(V; X_1^n)$$

which gives us a corresponding lower bound on $\mathfrak{M}_n(\mathcal{P}, \rho, \epsilon, \delta)$.

5 Non-Homogeneous Local Differential Privacy

As discussed in the previous section, the analysis of LDP assumes that all samples are privatized (whether interactively or non-interactively) with the same privacy parameters (ϵ, δ). A curious mind may ask, what if instead of privatizing all samples equally we do a non-homogeneous privatization. Such a scenario may occur, for example, if the database holds data from users who requir or demand different levels of privatization. Another possibility is as follows: suppose the users are part of a network where their generated data travels to the central database via relay nodes, each of which applies its own privatization mechanism. Thus resulting in differently privatized samples at the database. This results in some samples having "low" privacy and some having "high" privacy. The information contraction perspective tells us that low privacy implies less contraction and hence more information is carried by these fraction of samples. One then asks how much better can the estimator do. With this regard we introduce the following generalized definition.

Definition 4. *For n i.i.d. samples generated from some underlying distribution over sample space $\mathcal{X}$, positive integer $t \leq n$, $\overline{\epsilon} \in \mathbb{R}_+^t, \overline{\delta} \in [0,1]^t, \overline{\rho} \in [0,1]^t, \sum_{i=1}^t \rho_i = 1$, a non-homogeneous LDP mechanism $\{\mathcal{K}_i\}_{i=1}^n$ is called $(\overline{\epsilon}, \overline{\delta}, \overline{\rho})$-LDP if ρ_i fraction of the total number of samples are (ϵ_i, δ_i) locally differentially private.*

Treating the vector $\overline{\rho}$ as a probability vector, one can hence guarantee an expected or average privacy level of the whole process. Below we introduce a model for such a mechanism.

Consider a finite tree $\mathcal{T} = (V, E)$. Starting at the root, every vertex generates m children for $m \in \mathbb{N}$ and among these m children a randomly selected fraction $\mu \in (0, 1)$ of them do not produce any further offspring (μ is chosen so as to make $m\mu$ an integer) and the other children further produce m offspring of their own. We group the vertices by layers in a natural way namely vertices at a distance i from the root are the layer i vertices. We stop this process at some layer t, i.e., non of the vertices in layer t reproduce any further. Let the set of all leaves of all layers of the tree generated this way be $\mathcal{L}$ and the leaves at layer i be $\mathcal{L}_i$ with $|\mathcal{L}_i| = l_i$. The root will generate $n = |\mathcal{L}|$ i.i.d. samples $\{X_j\}_{j=1}^n$ from some underlying distribution and each sample is mapped to one of the leaves by a uniformly chosen permutation $\pi_n : [n] \to \mathcal{L}$. Let $\mathcal{K} : \mathcal{X} \to \mathcal{X}$ be an (ϵ, δ) LDP channel. We denote the i-fold composition of the channel $\mathcal{K}$ by $\mathcal{K}^i$. A sample X_j is privatized through $\mathcal{K}^i$ where i is the distance of $\pi_n(j)$ from the root. Another way to think of this

mechanism is that each sample generated at the root "travels" to its corresponding leaf vertex via the unique path and each edge on that path acts independently as a privatizing channel $\mathcal{K}$.

Claim 3. *For reasonably chosen m, μ, t and $\mathcal{K} \in \mathfrak{K}_{\epsilon,\delta}$, the above mechanism satisfies $(\overline{\epsilon}, \overline{\delta}, \overline{\rho})$-LDP with $\epsilon_i = \epsilon,\ \delta_i = \delta^i,\ \ \forall\ i \in [t]$ and $\rho_i \approx \frac{\mu}{m^{t-i-1}(1-\mu)^{t-i-1}(\mu+m-m\mu)}\ \ \forall\ i \in [t-1], \rho_t \approx \frac{m-m\mu}{\mu+m-m\mu}$.*

Proof. At any layer of the t-layer tree, among the m children of any vertex, $(1-\mu)m$ children further reproduce m children and μm children become leaves at that level. A careful but straightforward calculation reveals that the number of leaves at layer i, $i \in [t-1]$ is $l_i = (1-\mu)^{i-1}m^i\mu$ and $l_t = (1-\mu)^{t-1}m^t$. Hence we have

$$\begin{aligned} n &= \sum_{i-1}^{t-1}(1-\mu)^{i-1}m^i\mu + (1-\mu)^{t-1}m^t \\ &= \frac{\mu}{(1-\mu)}\frac{m^t(1-\mu)^t - 1}{m(1-\mu)-1} + (1-\mu)^{t-1}m^t \\ &\approx m^{t-1}(1-\mu)^{t-2}\mu + (1-\mu)^{t-1}m^t \\ &= m^{t-1}(1-\mu)^{t-2}(\mu + m - m\mu) \end{aligned}$$

And so

$$\rho_i = \frac{l_i}{n} \approx \frac{(1-\mu)^{i-1}m^i\mu}{m^{t-1}(1-\mu)^{t-2}(\mu+m-m\mu)} = \frac{\mu}{m^{t-i-1}(1-\mu)^{t-i-1}(\mu+m-m\mu)} \quad 1 \le i \le t-1$$

and

$$\rho_t = \frac{l_t}{n} = \frac{m-m\mu}{\mu+m-m\mu}$$

For the ϵ_i and δ_i, we need to characterize the equivalent privacy guarantees of the i-fold composition channel $\mathcal{K}^i$. For this we resort to Theorem 1 and use it i times for two distributions P, Q on $\mathcal{X}$ to get

$$\begin{aligned} E_{e^\epsilon}(P\mathcal{K}^i||Q\mathcal{K}^i) &\le \delta E_{e^\epsilon}(P\mathcal{K}^{i-1}||Q\mathcal{K}^{i-1}) \\ &\le \cdots \\ &\le \delta^i E_{e^\epsilon}(P||Q) \end{aligned}$$

which implies by the equivalence theorem that $\mathcal{K}^i$ is an (ϵ, δ^i)-LDP channel. This completes the proof of the claim. □

Remark: An alternative way to generate such a tree for a given n and proportion vector $\overline{\rho}$ would be as follows: Take a regular tree of degree m and sufficient depth t and select uniformly randomly from the nodes of layer i a number of $n\rho_i$ of nodes for each i. The travel paths of each sample might not be unique in this case anymore but the privacy guarantees and analysis will still hold.

5.1 Non-homogeneous Locally Private La Cam's Method

For the tree-based privatization mechanism described above, let M_i^n be the induced output distribution of Y_1^n when the inputs are generated from P_i, $i \in \{0,1\}$. Notice that since the input distribution of X_1^n is the product distribution $P_i^{\otimes n}$ and the privatization mechanisms of each sample are independent (although not identical), M_i^n is also a product distribution. But

the individual mechanisms are different namely l_1 number of samples $X_{\pi_n^{-1}(\mathcal{L}_1)}$ go through the channel $\mathcal{K}^{\otimes l_1}$, l_2 number of samples $X_{\pi_n^{-1}(\mathcal{L}_2)}$ go through the two fold composition of $\mathcal{K}^{\otimes l_2}$, i.e., $(\mathcal{K}^{\otimes l_2})^2 \equiv (\mathcal{K}^2)^{\otimes l_2}$ and so on. We have the following lemma.

Lemma 4.

$$D_{TV}^2(M_0^n||M_1^n) \leq \frac{1}{2}\sum_{i=1}^{t} \phi_{l_i}(\epsilon,\delta)^i l_i D_{KL}(P_0||P_1)$$

Proof.

$$\begin{aligned}
&D_{TV}^2(((\mathcal{K})^{\otimes l_1} \times \cdots \times (\mathcal{K}^t)^{\otimes l_t})P_0^{\otimes n}||((\mathcal{K})^{\otimes l_1} \times \cdots \times (\mathcal{K}^t)^{\otimes l_t})P_1^{\otimes n}) \\
&\overset{(1)}{\leq} \frac{1}{2} D_{KL}(((\mathcal{K})^{\otimes l_1} \times \cdots \times (\mathcal{K}^t)^{\otimes l_t})P_0^{\otimes n}||((\mathcal{K})^{\otimes l_1} \times \cdots \times (\mathcal{K}^t)^{\otimes l_t})P_1^{\otimes n}) \\
&\overset{(2)}{=} \frac{1}{2}\sum_{i=1}^{t} D_{KL}((\mathcal{K}^i)^{\otimes l_i}P_0^{\otimes l_i}||(\mathcal{K}^i)^{\otimes l_i}P_1^{\otimes l_i}) \\
&\overset{(3)}{\leq} \frac{1}{2}\sum_{i=1}^{t} \eta_{KL}(\mathcal{K}^{\otimes l_i}) D_{KL}((\mathcal{K}^{i-1})^{\otimes l_i}P_0^{\otimes l_i}||(\mathcal{K}^{i-1})^{\otimes l_i}P_1^{\otimes l_i}) \\
&\overset{(3)}{\leq} \cdots \\
&\overset{(3)}{\leq} \frac{1}{2}\sum_{i=1}^{t} \eta_{KL}(\mathcal{K}^{\otimes l_i})^i D_{KL}(P_0^{\otimes l_i}||P_1^{\otimes l_i}) \\
&\overset{(4)}{=} \frac{1}{2}\sum_{i=1}^{t} \phi_{l_i}(\epsilon,\delta)^i l_i D_{KL}(P_0||P_1)
\end{aligned}$$

where (1) follows from Pinsker's inequality, (2) and (4) follow from the chain rule of divergence and (3) follows from the successive use of Equation 4. □

With this we have the following theorem.

Theorem 5. *For the privatization mechanism based upon the tree $T(m,\mu,t)$, the minmax estimation error is lower bounded as follows*

$$\mathfrak{M}_n(\mathcal{P},\rho,\epsilon,\delta,m,\mu,t) \geq \frac{\tau}{2}\left[1 - \frac{1}{\sqrt{2}}\sqrt{\sum_{i=1}^{t} \phi_{l_i}(\epsilon,\delta)^i l_i D_{KL}(P_0||P_1)}\right]$$

Proof. Follows directly from 4. □

5.2 Non-homogeneous Locally Private Fano's Method

Recall Section 4.2 where a lower bound on the estimation error was found by finding an upper bound on $I(V;Y_1^n)$ and then applying Fano's inequality. The following lemma gives a similar upper bound on $I(V;Y_1^n)$ for the non-homogeneous privatization scheme.

Lemma 6.

$$I(V;Y_1^n) \leq \sum_{i=1}^{t} \phi_{l_i}(\epsilon,\delta)^i I(V;X_{\pi_n^{-1}(\mathcal{L}_i)})$$

Proof. We denote $Y_{\mathcal{L}_i}$ to be the set of privatized samples at layer i. Then we have,

$$
\begin{aligned}
I(V;Y_1^n) &= I(V;Y_{\mathcal{L}_1},Y_{\mathcal{L}_2},\cdots,Y_{\mathcal{L}_t}) \\
&\overset{(1)}{\leq} I(V;Y_{\mathcal{L}_1}) + I(V;Y_{\mathcal{L}_2}|Y_{\mathcal{L}_1}) + \cdots + I(V;;Y_{\mathcal{L}_t}|Y_{\mathcal{L}_1},Y_{\mathcal{L}_2}\cdots,Y_{\mathcal{L}_{t-1}}) \\
&= I(V;Y_{\mathcal{L}_1}) + H(Y_{\mathcal{L}_2}|Y_{\mathcal{L}_1}) - H(Y_{\mathcal{L}_2}|V,Y_{\mathcal{L}_1}) + \cdots + I(V;;Y_{\mathcal{L}_t}|Y_{\mathcal{L}_1},Y_{\mathcal{L}_2}\cdots,Y_{\mathcal{L}_{t-1}}) \\
&\overset{(2)}{=} I(V;Y_{\mathcal{L}_1}) + H(Y_{\mathcal{L}_2}|Y_{\mathcal{L}_1}) - H(Y_{\mathcal{L}_2}|V) + \cdots + I(V;;Y_{\mathcal{L}_t}|Y_{\mathcal{L}_1},Y_{\mathcal{L}_2}\cdots,Y_{\mathcal{L}_{t-1}}) \\
&\overset{(3)}{\leq} I(V;Y_{\mathcal{L}_1}) + H(Y_{\mathcal{L}_2}) - H(Y_{\mathcal{L}_2}|V) + \cdots + I(V;;Y_{\mathcal{L}_t}|Y_{\mathcal{L}_1},Y_{\mathcal{L}_2}\cdots,Y_{\mathcal{L}_{t-1}}) \\
&= I(V;Y_{\mathcal{L}_1}) + I(V;Y_{\mathcal{L}_2}) + H(Y_{\mathcal{L}_3}|Y_{\mathcal{L}_2},Y_{\mathcal{L}_1}) - H(Y_{\mathcal{L}_3}|V,Y_{\mathcal{L}_2},Y_{\mathcal{L}_1}) + \cdots, + I(V;Y_{\mathcal{L}_t}|Y_{\mathcal{L}_1},Y_{\mathcal{L}_2}\cdots,Y_{\mathcal{L}_{t-1}}) \\
&\overset{(2),(3)}{\leq} \cdots \\
&\overset{(2),(3)}{\leq} \sum_{i=1}^{t} I(V;Y_{\mathcal{L}_i}) \\
&\overset{(4)}{\leq} \sum_{i=1}^{t} \phi_{l_i}(\epsilon,\delta)^i I(V;X_{\pi_n^{-1}(\mathcal{L}_i)})
\end{aligned}
$$

where (1) follows from chain rule, (2) follows because $Y_{\mathcal{L}_i} \leftrightarrow V \leftrightarrow Y_{\mathcal{L}_j}$ forms a Markov chain for $i \neq j$, (3) follows because conditioning reduces entropy and (4) follows due to successive application of Equation 4. □

Theorem 7. *For the privatization mechanism based upon the tree $T(m,\mu,t)$, the minmax estimation error is lower bounded as follows*

$$\mathfrak{M}_n(\mathcal{P},\rho,\epsilon,\delta,m,\mu,t) \geq \tau\left[1 - \frac{\sum_{i=1}^{t}\phi_{l_i}(\epsilon,\delta)^i I(V;X_{\pi_n^{-1}(\mathcal{L}_i)}) + \log 2}{\log|\mathcal{V}|}\right]$$

Proof. Follows directly from 6. □

6 Conclusion and Future Directions

In this work, an attempt has been made to understand this newly established strong connections between LDP and the well studied field of contraction of divergences. It is clear that these connections provide an alternative, perhaps easy to understand perspective to the field of privacy and vastly simplify and streamline the analysis of certain problems. Additionally, the topic of non-homogeneous local differential privacy invite several interesting questions of its own. Perhaps the first pertinent question would be to see how tight these bounds are that have been established in Section 5. Namely, what kind of schemes can be designed to achieve these bounds while satisfying the desired privacy guarantees. For example, if we assume binary alphabets and binary symmetric channels to be the privatization mechanism in the tree based scheme, what error rates can one achieve.

It is obvious that the techniques used to prove the bounds are not limited to the tree based scheme and can be applied to any such non-homogeneous privatization mechanism as long as it is well defined and there are certain independence properties. This opens up possibilities for several imaginative schemes. One of which, perhaps having reasonable theoretical and also practical interest, is to have a random tree instead of a deterministic one and study the "most probable" behavior with probabilistic guarantees.

References

[1] R. Ahlswede and P. Gacs. Spreading of Sets in Product Spaces and Hypercontraction of the Markov Operator. *The Annals of Probability*, 4(6):925 – 939, 1976. 4

[2] S. Asoodeh, M. Aliakbarpour, and F. P. Calmon. Local differential privacy is equivalent to contraction of e_γ-divergence. 2021. 3, 4, 5, 6

[3] S. Asoodeh, M. Diaz, and F. P. Calmon. Privacy analysis of online learning algorithms via contraction coefficients, 2020. 4

[4] J. E. Cohen, J. Kemperman, and G. Zbăganu. Comparisons of stochastic matrices, with applications in information theory, statistics, economics, and population sciences. *Journal of the American Statistical Association*, 94:984, 1999. 4

[5] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy, data processing inequalities, and statistical minimax rates, 2014. 4

[6] C. Dwork, G. N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60, 2010. 4