# Local Differential Privacy and Strong Data Processing Inequalities

## ENEE729T: Information theoretic methods in Learning

Adway Patra

September 19, 2021

University of Maryland, College Park

# Table of contents

# Intro

# Introduction

- A major challenge in today's age of big data and machine learning is to balance statistical efficiency with user privacy.
- Differential privacy (DP) has become a standard definition for designing large-scale privacypreserving algorithms in both industrial and academic settings.
- The basic idea is that any curious accessor of the database should not be able to infer much about any particular user's entry by making queries.

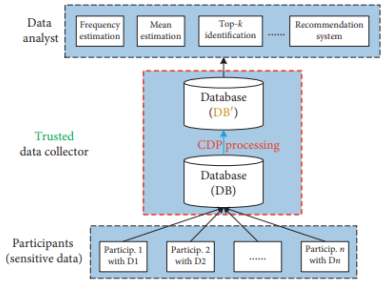
**Differential Privacy**

Randomized mechanism $\mathcal{K} : \mathcal{X}^n \to \mathcal{Y}$, is $(\epsilon, \delta)$-DP,

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

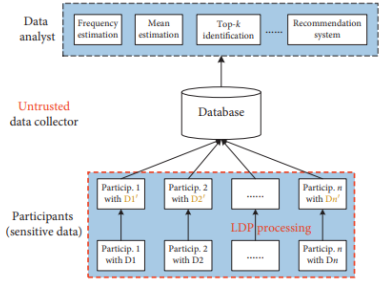
# Introduction: Centralized vs Local

## Differential Privacy




- Adds noise to the output query.

## Local Differential Privacy




- Adds noise to each sample in the database.

### Local Differential Privacy

Randomized mechanism $\mathcal{K} : \mathcal{X} \to \mathcal{Y}$ is $(\epsilon, \delta)$ LDP, iff

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

[1] Images taken from [5]

# Introduction: Strong Data Processing Inequalities

- 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.

- - $f(x) := x \log x \to$ KL- divergence.
  - $f(x) := \frac{1}{2}|x - 1| \to$ Total Variation Distance.
  - $f(x) := (x - 1)^2 \to \chi^2$-divergence.
  - $f_\gamma(x) := \max\{x - \gamma, 0\} \to E_\gamma$- divergence.

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

# Introduction: Strong Data Processing Inequalities..continued

Multiplication of a componentwise non-negative vector by a stochastic matrix results in a vector that is "more uniform".

- For transition kernel $K :\rightarrow$

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

- Inequality often strick: leads to strong data processing inequalities.
- Strictness measured in terms of contraction coefficient

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

- For the popular divergences, $\eta_{TV}(\mathcal{K}), \eta_{KL}(\mathcal{K}), \eta_{\chi^2}(\mathcal{K})$ and $\eta_\gamma(\mathcal{K})$ follow similar definitions.

[2] For more on SDPIs, refer [4]

# Equivalence between LDP and SDPI

# Equivalence between LDP and SDPI

$$\mathcal{K} \text{ is } (\epsilon, \delta) - LDP \iff E_{e^\epsilon}(P\mathcal{K}||Q\mathcal{K}) \leq \delta E_{e^\epsilon}(P||Q)$$

Additionally,

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

$$\phi(\epsilon, \delta) = 1 - e^{-\epsilon}(1 - \delta)$$

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

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

[3]

[3] Proofs in [2]

# Applications: Private Estimation

# Applications to Private Estimation

- A set of probability distributions $\mathcal{P}$ on some alphabet $\mathcal{X}$
- We wish to estimate a functional $\theta : \mathcal{P} \to \Theta$
- Normal Setting: Have access to i.i.d. samples $X_1^n$. Find estimator $\hat{\theta} : \mathcal{X}^n \to \Theta$
- Performance measured in terms of loss function $\rho : \Theta \times \Theta \to \mathbb{R}_+$
- Minmax error:

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

# Applications to Private Estimation

- In private setting, samples $X_1^n$ are privatized by mechanisms $\{\mathcal{K}_i\}_{i=1}^n$, each $(\epsilon, \delta)$ private. The outputs $Y_1^n$ are revealed.
- $\{\mathcal{K}_i\}_{i=1}^n$ can be interactive or non-interactive.
- Find estimator $\psi : \mathcal{Y}^n \to \Theta$
- Minmax error:

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

# Applications to Private Estimation: Le Cam's Method

- Applicable when differntiating between two distributions $P_0$ and $P_1$.
- Basically a binary hypothesis testing problem.
- Uses TV distance to bound the error:

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

- Minmax error (Non-private Setting):

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

- In the private setting, $D_{TV}(P_0^{\otimes n}||P_1^{\otimes n})$ is replaced by the TV distance of the two induced output distributions.
- Minmax error (Private Setting):

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

[4]

[4] Results taken from [3],[2]

# Applications to Private Estimation: Fano's Method

- When there is an indexed finite set of distributions.
- Bounds using mutual information between the index random variable $V$ and the samples.
- Minmax error (Non-private Setting):

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

- In private setting, replace $I(V; X_1^n)$ with $I(V; Y_1^n)$. Use following bound:

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

5

[5] Results taken from [3],[2]

# Non-homogeneous Privacy

# Non-homogeneous Privacy : Motivation

- Local privacy definitions usually employ same $(\epsilon, \delta)$ for all samples, whether mechanism is interactive or non-interactive.
- What if we want different privacy levels for different samples.
- Some data entries might be more sensitive than others.
- Another scenario: Suppose each individual sends their data through a network of relays and each relay performs and independent privatization. Different path lengths will translate to different privacy levels.

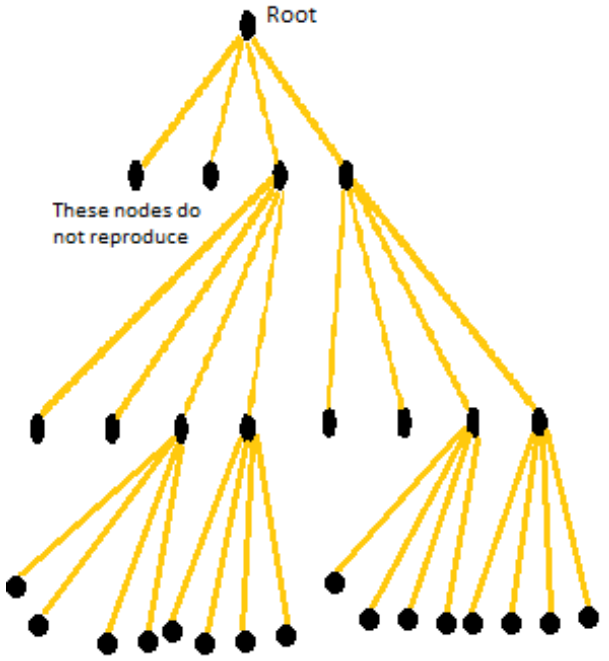
**Non-homogeneous LDP**

For $X_1^n \in \mathcal{X}^n$ i.i.d. samples , 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 $\rho_i$ fraction of the total number of samples are $(\epsilon_i, \delta_i)$ locally differentially private.

Remark: The idea of the definition is similar to the definition of Heterogeneous Differential Privacy of [1]

# Non-homogeneous Privacy : A possible scheme

- Consider a finite tree $T = (V, E)$.
- Every vertex, starting from the root, creates $m$ children.
- $\mu$ fraction of those children do not reproduce further and the rest produce $m$ children of their own.
- Stop the tree at any level $t$, i.e., all nodes in layer $t$ stop reproducing further.
- Root produces i.i.d. samples.
- Each edge acts independently as a $(\epsilon, \delta)$ LDP channel.
- Each sample travels to one of the leaves.
- Result: Due to different path lengths, samples are non-homogeneously privatized.



# Non-homogeneous Privacy : Analysis

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

$l_i = (1-\mu)^{i-1} m^i \mu$ and $l_t = (1-\mu)^{t-1} m^t$.

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

For the $i$-fold composition channel [i]

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

# Non-homogeneous Privacy : Le Cam's Method

## Recall

$$P_{err}(V|Y_1^n) \geq \frac{1}{2}[1 - D_{TV}(M_0^n||M_1^n)]$$

So we need bound on $D_{TV}(M_0^n||M_1^n)$.

## Theorem 1

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

- Use Pinsker's inequality to convert to KL-divergence.
- Use independence property to separate each layer.
- Use iteratively for each layer

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

# Non-homogeneous Privacy : Fano's Method

## Recall

We need upper bound on $I(V; Y_1^n)$

## Theorem 2

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

- Use chain rule.
- Separate layers using the fact that conditioned on $V$, $Y_{\mathcal{L}_i}$ is independent of $Y_{\mathcal{L}_j}$.
- Use the mutual information contraction iteratively for each layer as before.

# Conclusion and Future work

# Conclusion and Future work

- In the tree-based model, due to geometrical growth, most samples have high privacy.
- Since some samples carry more information, the performance should be better. How better is the performance from homogeneous case?
- Techniques used not limited to the particular tree-based privatization model. What other models possible?
- What if the tree was random?

Simplest thing to imagine: $P_0$ and $P_1$ be Bernoulli and each channel be $(\epsilon, \delta)$ LDP Binary Symmetric Channels, aka, the Randomized Response mechanism. How better is the performance?

# References i


- M. Alaggan, S. Gambs, and A.-M. Kermarrec.
  **Heterogeneous dierential privacy.**
- S. Asoodeh, M. Aliakbarpour, and F. P. Calmon.
  **Local differential privacy is equivalent to contraction of $e_\gamma$-divergence, 2021.**
- J. Duchi, M. Wainwright, and M. Jordan.
  **Minimax optimal procedures for locally private estimation, 2017.**
- Y. Polyanskiy and Y. Wu.
  **Strong data-processing inequalities for channels and bayesian networks, 2016.**

# References ii


X. Xiong, S. Liu, D. Li, Z. Cai, and X. Niu.
**A comprehensive survey on local differential privacy.**
*Security and Communication Networks, Article ID 8829523*, 2020:403–422, 2020.

Questions?