

Exposé de candidature MCF pour le poste N°4887  
..en Statistique et Thèmes Connexes (Section 26) à l'UT3 Paul Sabatier / l'IMT

Clément Lalanne

Chercheur postdoctoral à Toulouse School of Economics

*clement.lalanne@tse-fr.eu*



2024-04-29

# Table of Contents

Introduction

Recherche

Enseignement

Conclusion

### 3 Curriculum

## 2016-2020 : Élève fonctionnaire stagiaire à l'ENS en Informatique



- ▶ Master 1 MPRI Master 2 MVA
- ▶ Stage de L3 : De-anonymization and privacy : Study of a random graph model encadré par Florian Simatos
- ▶ Stage de M1 : Storage-optimal continuous optimization encadré par Volkan Cevher
- ▶ Stage de M2 : Nystagmus waveform extraction using convolutional dictionary learning with detrending encadré par Laurent Oudre, Nicolas Vayatis et Thomas Moreau



### 2020-2023 : Doctorat à l'ENS de Lyon en Informatique



- ▶ Sujet : **Sur les compromis liés à l'apprentissage statistique sous contraintes de confidentialité**
- ▶ Encadré par **Aurélien Garivier** et **Rémi Gribonval**
- ▶ Chargé de TDs (3x64h) :
  - ▶ Apprentissage Statistique (M1 ENS Lyon)
  - ▶ Bases de données et Exploration de données (M1 ENS Lyon)
  - ▶ Entrainement à la programmation sportive (L3 ENS Lyon)
  - ▶ Optimisation (M1 ENS Lyon)
  - ▶ Programmation Système (IUT 2A UCBL)
  - ▶ Évaluation de stages, ...



### 2023-Ajd : Postdoc en Mathématiques à TSE



- ▶ Sujet : **Optimisation des réseaux de neurones profonds**
- ▶ Encadré par **Jérôme Bolte et Sébastien Gadat**



# Table of Contents

Introduction

**Recherche**

Enseignement

Conclusion

## 7 Thèmes de recherche et Publications

### Thèmes de recherche : Statistique confidentielle, Optimisation

#### Preprints :

- ▶ ★ **Privately Learning Smooth Distributions on the Hypercube by Projections**, Clément Lalanne et Sébastien Gadat, 2024

#### Journaux :

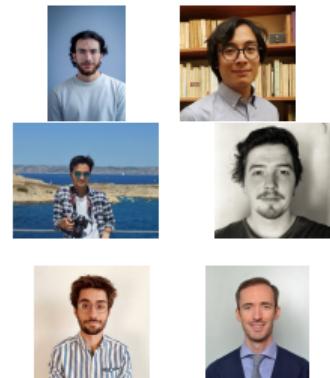
- ▶ ★ **About the Cost of Central Privacy in Density Estimation**, Clément Lalanne, Aurélien Garivier et Rémi Gribonval, dans **Transactions on Machine Learning Research**, 2023
- ▶ **Private Quantiles Estimation in the Presence of Atoms**, Clément Lalanne, Clément Gastaud, Nicolas Grislain, Aurélien Garivier et Rémi Gribonval, dans **Information and Inference**, 2023
- ▶ ★ **On the Statistical Complexity of Estimation and Testing under Privacy Constraints**, Clément Lalanne, Aurélien Garivier et Rémi Gribonval, dans **Transactions on Machine Learning Research**, 2023



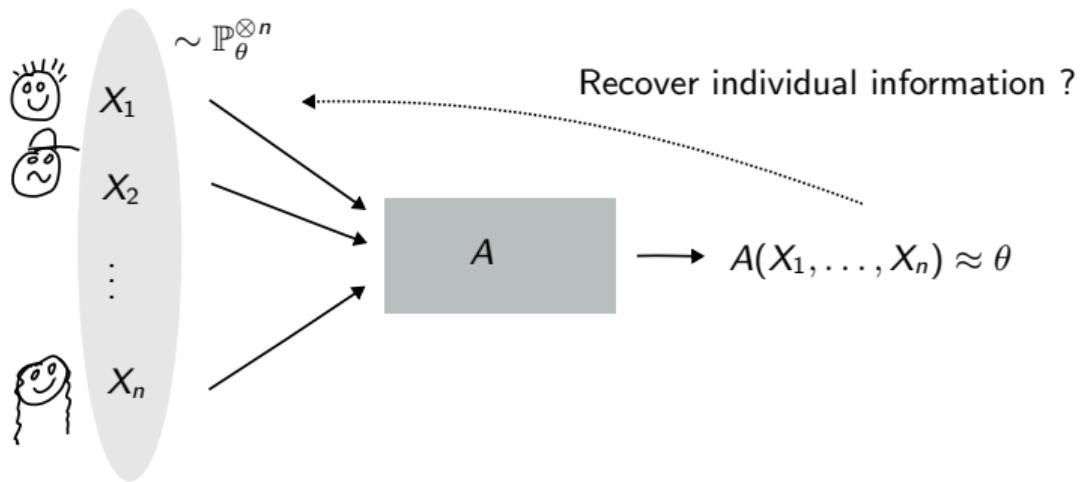
## 8 Publications

### Conférences :

- ▶ **Private Statistical Estimation of Many Quantiles,**  
Clément Lalanne, Aurélien Garivier et Rémi Gribonval, dans  
**International Conference on Machine Learning**, 2023
- ▶ **Can sparsity improve the privacy of neural networks?,**  
Antoine Gonon, Léon Zheng, Clément Lalanne, Quoc-Tung Le,  
Guillaume Lauga et Can Pouliquen, dans **GRETSI** (francophone),  
2023
- ▶ **Extraction of Nystagmus Patterns from Eye-Tracker  
Data with Convolutional Sparse Coding**, Clément Lalanne,  
Maxence Rateaux, Laurent Oudre, Matthieu Robert et Thomas  
Moreau, dans **IEEE Engineering in Medicine and Biology  
Society**, 2020



## 9 Differential Privacy 101



**Question :**

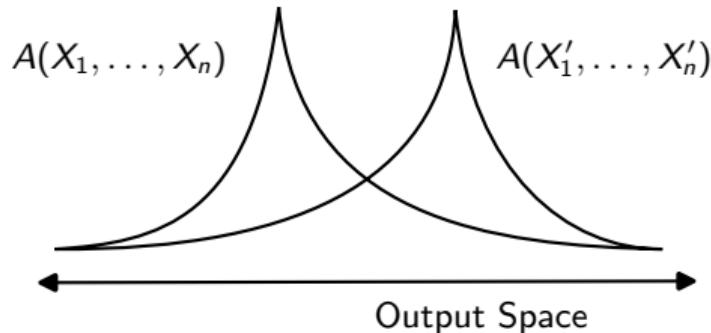
*Peut-on se défendre contre toute attaque ?*

## 10 Differential Privacy 101

**Voisinage :**  $\mathbf{X} \sim \mathbf{X}'$  ssi  $\mathbf{X}$  peut être obtenu à partir de  $\mathbf{X}'$  en changeant les données d'un individu.

**Définition :**  $A$  est  $\epsilon > 0$ -DP si  $A(\mathbf{X})$  est  $\epsilon$ -proche de  $A(\mathbf{X}')$  pour tous  $\mathbf{X} \sim \mathbf{X}'$ .<sup>1</sup>

**Rôle de  $\epsilon$  :** Plus  $\epsilon$  est petit, plus  $A$  est confidentiel.



**Question :**

*Quel est le coût de la confidentialité sur l'estimation statistique ?*

---

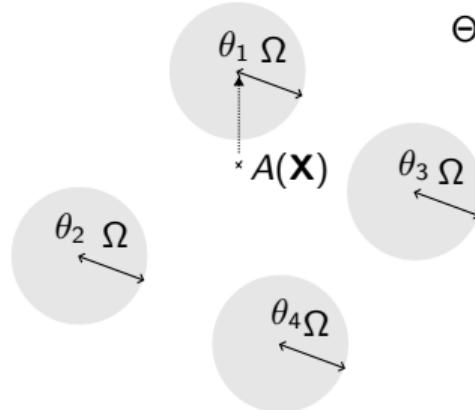
<sup>1</sup>Cynthia Dwork et al. "Calibrating Noise to Sensitivity in Private Data Analysis". In: 2006.

## 11 Un schéma de preuve pour les bornes inférieures

**Setup :**  $\theta \in \Theta$ ,  $\mathbf{X} = (X_1, \dots, X_n) \sim \mathbb{P}_\theta^{\otimes n}$ .

**Risque minimax :**<sup>23</sup>

$$\begin{aligned}\mathfrak{M}_n := \inf_{A} \sup_{\theta \in \Theta} \mathbb{E}_{A, \mathbf{X} \sim \theta} (\text{Error}(A(\mathbf{X}), \theta)) &\geq \inf_A \sup_{i=1, \dots, N} \mathbb{E}_{A, \mathbf{X} \sim \theta_i} (\text{Error}(A(\mathbf{X}), \theta_i)) \\ &\geq \Phi(\Omega) \sup_{i=1, \dots, N} \mathbb{P}_{A, \mathbf{X} \sim \theta_i} (\hat{i}(A(\mathbf{X})) \neq i)\end{aligned}$$



<sup>2</sup>Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. 2009.

<sup>3</sup>Error(·, ·) = Φ(d(·, ·))

## 12 Un schéma de preuve pour les bornes inférieures

**Risque minimax :**

$$\mathfrak{M}_n \geq \Phi(\Omega) \sup_{i=1, \dots, N} \mathbb{P}_{A, \mathbf{X} \sim \theta_i} \left( \hat{i}(A(\mathbf{X})) \neq i \right)$$

**Problème de transport :**

$$\sup_{i=1, \dots, N} \mathbb{P}_{A, \mathbf{X} \sim \theta_i} \left( \hat{i}(A(\mathbf{X})) \neq i \right) \geq \boxed{\sup_{\mathbb{Q} \in \Pi(\mathbb{P}_1^{\otimes n}, \dots, \mathbb{P}_N^{\otimes n})} \int s(\mathbf{X}_1, \dots, \mathbf{X}_N) d\mathbb{Q}(\mathbf{X}_1, \dots, \mathbf{X}_N)},$$

where  $s$  is a *similarity function* satisfying, for any  $\mathbf{X}_1, \dots, \mathbf{X}_N$ ,

$$\boxed{\frac{1}{N} \sum_{i=1}^N \mathbb{P}_A \left( \hat{i}(A(\mathbf{X}_i)) \neq i \right) \geq s(\mathbf{X}_1, \dots, \mathbf{X}_N)}.$$

## 13 Un schéma de preuve pour les bornes inférieures

**Risque minimax :**

$$\mathfrak{M}_n \geq \Phi(\Omega) \sup_{i=1, \dots, N} \mathbb{P}_{A, \mathbf{x} \sim \theta_i} \left( \hat{i}(A(\mathbf{X})) \neq i \right)$$

**Le Cam :**<sup>4</sup>

$$\boxed{\sup_{i=1,2} \mathbb{P}_{A, \mathbf{x} \sim \theta_i} \left( \hat{i}(A(\mathbf{X})) \neq i \right) \geq \frac{1}{2} \left( 1 - (1 - e^{-\epsilon}) \text{TV}(\mathbb{P}_{\theta_1}, \mathbb{P}_{\theta_2}) \right)^n}.$$

**Fano :**<sup>5</sup>

$$\boxed{\sup_{i=1, \dots, N} \mathbb{P}_{A, \mathbf{x} \sim \theta_i} \left( \hat{i}(A(\mathbf{X})) \neq i \right) \geq 1 - \frac{1 + \frac{n\epsilon}{N^2} \sum_{i,j=1}^N \frac{2\text{TV}(\mathbb{P}_{\theta_i}, \mathbb{P}_{\theta_j})}{1 + \text{TV}(\mathbb{P}_{\theta_i}, \mathbb{P}_{\theta_j})}}{\ln(N)}}.$$

<sup>4</sup>  $\text{TV}(\mathbb{P}_1, \mathbb{P}_2) := \sup_S \mathbb{P}_1(S) - \mathbb{P}_2(S)$

<sup>5</sup>  $\text{KL}(\mathbb{P}_1 \| \mathbb{P}_2) := \int \ln \left( \frac{d\mathbb{P}_1}{d\mathbb{P}_2} \right) d\mathbb{P}_1$

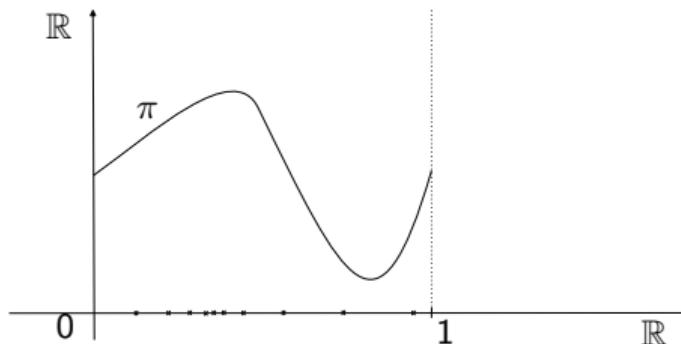
## 14 Application à l'estimation de densités

**Vitesse d'estimation :**  $\Theta\left(n^{-\frac{2\beta}{2\beta+d}} + (n\epsilon)^{-\frac{2\beta}{\beta+d}}\right)$

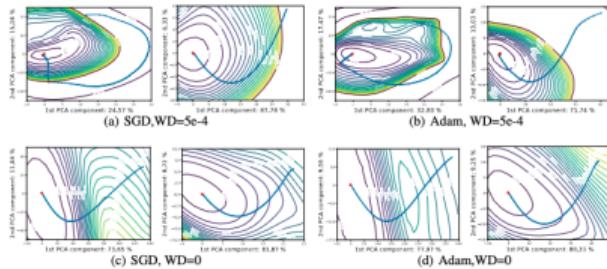
**Difficulté technique :** Pour les bornes inférieures, espaces de Sobolev-Hölder

$$\sum_{|\alpha|=\lfloor \beta \rfloor} \|\partial^\alpha \pi\|_2^2 + \mathbf{1}_{\beta-\lfloor \beta \rfloor > 0} \sum_{|\alpha|=\lfloor \beta \rfloor} \|\partial^\alpha \pi\|_{\mathcal{H}_{\beta-\lfloor \beta \rfloor}}^2 \leq L^2$$

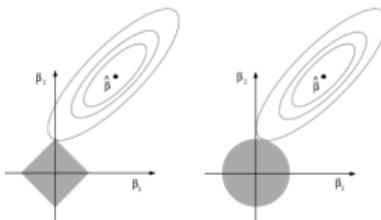
**Adaptivité :** Possible en adaptant la méthode de Lepskii + Polylog



### Optimisation des réseaux de neurones profonds par des méthodes du second ordre Travaux actuels de postdoc

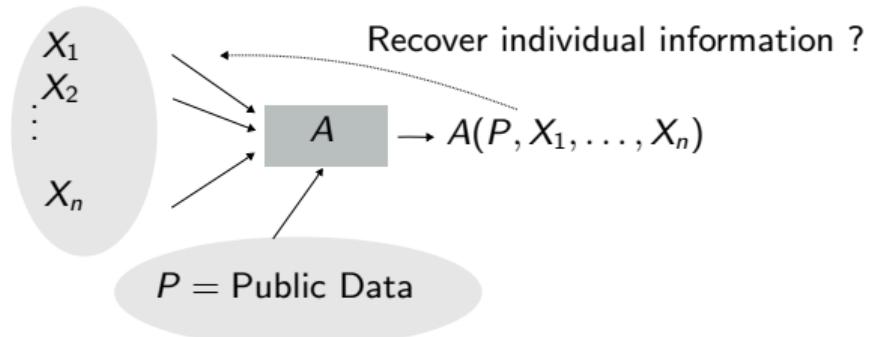


### Réduire le fléau de la dimension avec confidentialité

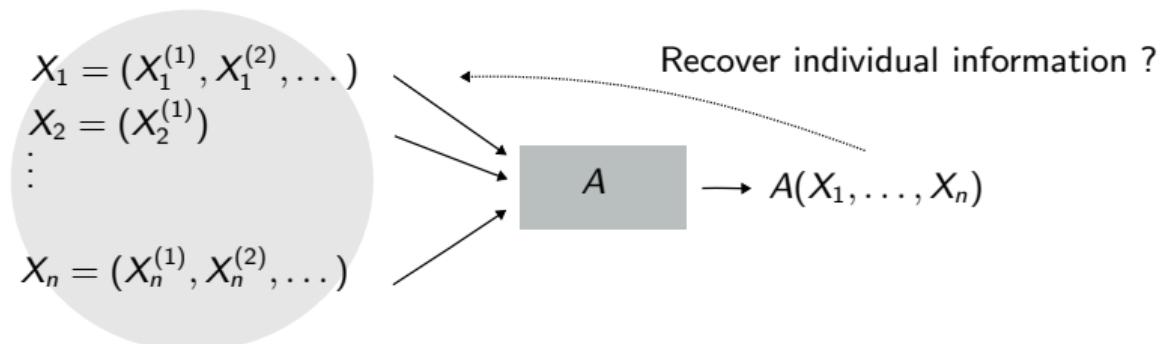


## 16 Projet de recherche

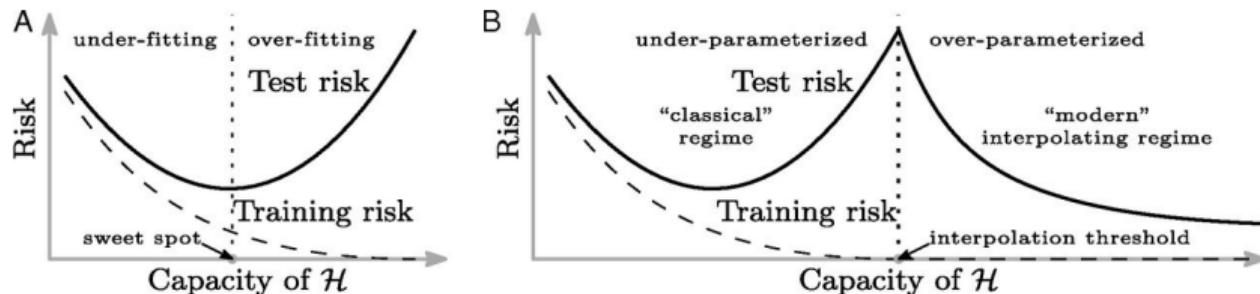
### Confidentialité mixte



### Confidentialité avec contributions asymétriques



### Confidentialité et sur-paramétrisation



### Temps d'entraînement des réseaux de neurones et confidentialité



#### Unlocking High-Accuracy Differentially Private Image Classification through Scale

Soham De<sup>\*1</sup>, Leonard Berrada<sup>\*1</sup>, Jamie Hayes<sup>1</sup>, Samuel L Smith<sup>1</sup> and Borja Balle<sup>1</sup>

<sup>\*</sup>Equal contributions, <sup>1</sup>DeepMind

### Projet de recherche proche des thématiques de l'équipe Statistiques et Optimisation de l'IMT

- ▶ Optimisation pour le machine learning
- ▶ Statistiques générales et statistiques non paramétriques
- ▶ Machine learning responsable (ex confidentialité, équité, robustesse)
- ▶ Possibles applications à des données médicales, aviation, ...

### Recherches d'actualité

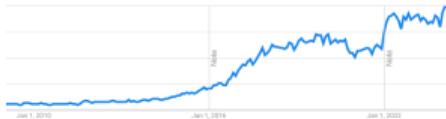


Figure: Google trend of "Machine Learning"

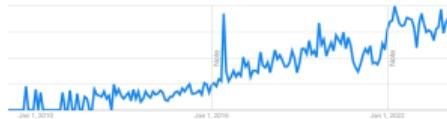


Figure: Google trend of "Differential Privacy"



Figure: Google trend of "GPU"

# Table of Contents

Introduction

Recherche

Enseignement

Conclusion

## 20 Expériences & qualifications

### Expérience : Chargé de TDs (3x64h)

- ▶ Apprentissage Statistique (M1 ENS Lyon)
- ▶ Bases de données et Exploration de données (M1 ENS Lyon)
- ▶ Entrainement à la programmation sportive (L3 ENS Lyon)
- ▶ Optimisation (M1 ENS Lyon)
- ▶ Programmation Système (IUT 2A UCBL)
- ▶ Évaluation de stages, ...

Agrégé de mathématiques (2023)

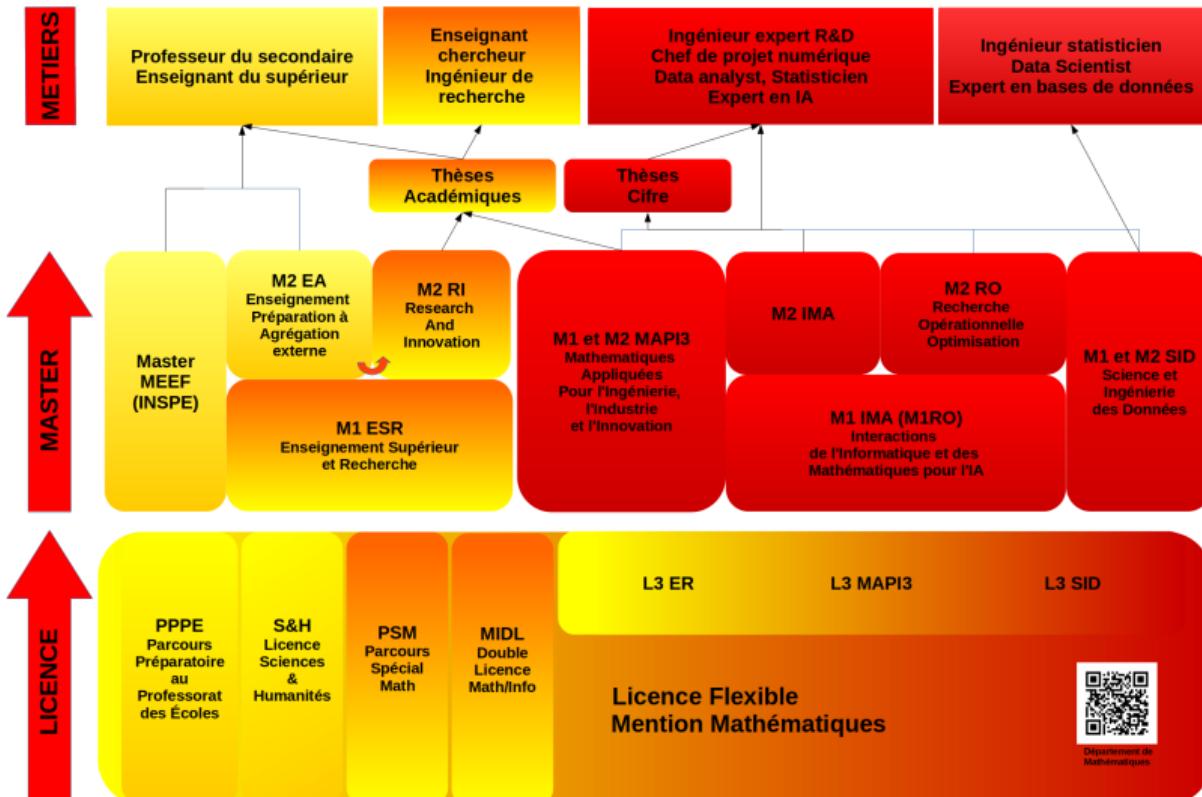
Maîtrise des outils techniques



Maîtrise de l'anglais

# 21 Intégration

## Offre de formation en Mathématiques 2022/2026



# Table of Contents

Introduction

Recherche

Enseignement

Conclusion

# 23 Pourquoi l'UT3 / l'IMT ?

## Excellence



## Écosystème de recherche



## Moyens



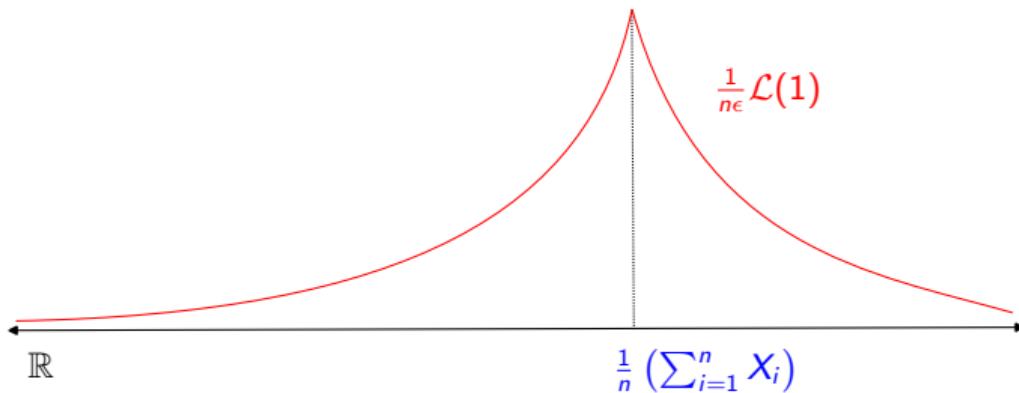
## Conclusion

**Merci de votre attention**

## 25 Private estimator

**Laplace mechanism :**

$$A(\mathbf{X}) = \frac{1}{n} \left( \sum_{i=1}^n X_i \right) + \frac{1}{n\epsilon} \mathcal{L}(1)$$



## 26 Error decomposition

**Laplace mechanism :**

$$A(\mathbf{X}) = \frac{1}{n} \left( \sum_{i=1}^n X_i \right) + \frac{1}{n\epsilon} \mathcal{L}(1)$$

is  $\epsilon$ -DP.

**Error :**

$$\mathbb{E}_{\mathbb{P}_{A,B(\theta)^{\otimes n}}} ((A(\mathbf{X}) - \theta))^2 \leq \frac{1/4}{n} + \frac{2}{n^2 \epsilon^2}$$

**Two regimes :**

- ▶ **Low privacy regime :**  $\epsilon = \Omega(1/\sqrt{n})$ , no significant effect on estimation.
- ▶ **High privacy regime :**  $\epsilon \ll 1/\sqrt{n}$ , the precision can be arbitrarily degraded.

**Question :**

*Is it possible to do better ?*

## 27 Building similarity functions

$$\frac{1}{N} \sum_{i=1}^N \mathbb{P}_A \left( \hat{i}(A(\mathbf{X}_i)) \neq i \right) \geq s(\mathbf{X}_1, \dots, \mathbf{X}_N).$$

**Definition :**  $A$  is  $\epsilon$ -DP if  $\mathbf{X} \sim \mathbf{X}' \implies \mathbb{P}(A(\mathbf{X}) \in S) \leq e^\epsilon \times \mathbb{P}(A(\mathbf{X}') \in S)$ .<sup>6</sup>

**Two marginals :**

$$\frac{1}{2} \sum_{i=1}^2 \mathbb{P}_A \left( \hat{i}(A(\mathbf{X}_i)) \neq i \right) \geq \frac{1}{2} e^{-\epsilon d_{\text{ham}}(\mathbf{X}_1, \mathbf{X}_2)} \quad ^7$$

**Many marginals :**

$$\frac{1}{N} \sum_{i=1}^N \mathbb{P}_A \left( \hat{i}(A(\mathbf{X}_i)) \neq i \right) \geq 1 - \frac{1 + \frac{\epsilon}{N^2} \sum_{i=1}^N \sum_{j=1}^N d_{\text{ham}}(\mathbf{X}_i, \mathbf{X}_j)}{\ln(N)}$$

<sup>6</sup>Dwork et al., "Calibrating Noise to Sensitivity in Private Data Analysis".

<sup>7</sup> $d_{\text{ham}}((X_1, \dots, X_n), (X'_1, \dots, X'_n)) := \sum_{i=1}^n \mathbb{1}_{X_i \neq X'_i}$

## 28 Back to the transport problem

$$\sup_{\mathbb{Q} \in \Pi(\mathbb{P}_1^{\otimes n}, \dots, \mathbb{P}_N^{\otimes n})} \int s(\mathbf{X}_1, \dots, \mathbf{X}_N) d\mathbb{Q}(\mathbf{X}_1, \dots, \mathbf{X}_N),$$

where  $s$  is **non-increasing** in  $d_{\text{ham}}(\mathbf{X}_i, \mathbf{X}_j)$  for any  $i, j$ .

**Question :**

*How to construct a coupling that makes those quantities big ?*

## 29 A good enough coupling

$$\sup_{\mathbb{Q} \in \Pi(\mathbb{P}_1^{\otimes n}, \dots, \mathbb{P}_N^{\otimes n})} \int s(\mathbf{X}_1, \dots, \mathbf{X}_N) d\mathbb{Q}(\mathbf{X}_1, \dots, \mathbf{X}_N),$$

where  $s$  is **non-increasing** in  $d_{\text{ham}}(\mathbf{X}_i, \mathbf{X}_j)$  for any  $i, j$ .

**Near optimal coupling for equalities** : There exists  $(X_i)_{i=1,\dots,N}$  of distribution  $\chi$ , a coupling between  $(\mathbb{P}_i)_{i=1,\dots,N}$  such that<sup>8</sup>

$$\forall i, j, \quad \text{TV}(\mathbb{P}_i, \mathbb{P}_j) \leq \boxed{\mathbb{P}(X_i \neq X_j) \leq \frac{2\text{TV}(\mathbb{P}_i, \mathbb{P}_j)}{1 + \text{TV}(\mathbb{P}_i, \mathbb{P}_j)}}.$$

**Final coupling** :  $\boxed{\mathbb{Q}^* = \chi^{\otimes n}}$

---

<sup>8</sup>Omer Angel and Yinon Spinka. *Pairwise optimal coupling of multiple random variables*. 2021.

## 30 $L_2$ approximations and projection estimators

Reference Fourier basis :

$$\phi_1(x) = 1$$

$$\phi_{2k}(x) = \sqrt{2} \sin(2\pi kx) \quad k \geq 1$$

$$\phi_{2k+1}(x) = \sqrt{2} \cos(2\pi kx) \quad k \geq 1 .$$

$L_2$  approximation :

$$\sum_{i=1}^N \theta_i \phi_i \xrightarrow[N \rightarrow +\infty]{L^2} \pi \quad \text{where} \quad \theta_i := \int_{[0,1]} \pi \phi_i .$$

Projection estimator :<sup>9</sup>

$$\hat{\pi}^{\text{proj}}(\mathbf{X}) = \sum_{i=1}^N \hat{\theta}_i \phi_i \quad \text{where} \quad \hat{\theta}_i := \frac{1}{n} \sum_{j=1}^n \phi_i(X_j) .$$

Question : *How do we add privacy ?*

---

<sup>9</sup>Tsybakov, *Introduction to Nonparametric Estimation*.

## 31 $L_2$ approximations and *private* projection estimators

**Private projection estimator :**<sup>10</sup>

$$\hat{\pi}^{\text{proj}}(\mathbf{X}) = \sum_{i=1}^N \left( \hat{\theta}_i + C_{\epsilon, N} \mathcal{L}(1) \right) \phi_i \quad \text{where} \quad \hat{\theta}_i := \frac{1}{n} \sum_{j=1}^n \phi_i(X_j).$$

$\hat{\pi}^{\text{proj}}$  is  $\epsilon$ -DP.

**Question :**

*What is the utility (error) of this estimator ?*

---

<sup>10</sup> Larry A. Wasserman and Shuheng Zhou. "A Statistical Framework for Differential Privacy". In: (2010).

## 32 Lower-bounds

**Lower-bound against  $\epsilon$ -DP estimators :**<sup>11</sup>

$$\inf_A \sup_{\pi} \mathbb{E} \left( \|A(\mathbf{X}) - \pi\|_{L_2}^2 \right) \geq C_{L,\beta} \max \left\{ n^{-\frac{2\beta}{2\beta+1}}, (n\epsilon)^{-\frac{2\beta}{\beta+1}} \right\}$$

**Best known upper-bound :**

$$\mathbb{E} \left( \|\hat{\pi}^{\text{proj}}(\mathbf{X}) - \pi\|_{L_2}^2 \right) \leq C_{L,\beta} \max \left\{ n^{-\frac{2\beta}{2\beta+1}}, (n\epsilon)^{-\frac{2\beta}{\beta+3/2}} \right\}$$

**Question :**

*Is it possible to bridge the gap ?*

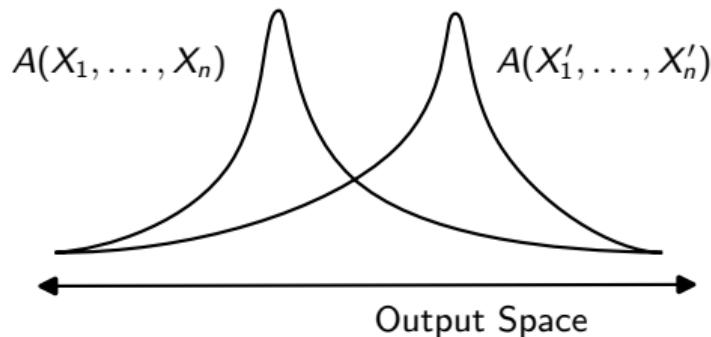
---

<sup>11</sup> Clément Lalanne, Aurélien Garivier, and Rémi Gribonval. “About the Cost of Central Privacy in Density Estimation”. In: (2023).

## 33 Concentrated Differential Privacy

**Definition :**<sup>12<sup>13</sup></sup>  $A$  is  $\rho$ -zCDP if  $\mathbf{X} \sim \mathbf{X}' \implies \forall \alpha > 0, D_\alpha(A(\mathbf{X}) \| A(\mathbf{X}')) \leq \alpha\rho$ , where

$$D_\alpha(\mathbb{P} \| \mathbb{Q}) := \frac{1}{\alpha - 1} \ln \int \left( \frac{d\mathbb{P}}{d\mathbb{Q}} \right)^{\alpha-1} d\mathbb{Q}.$$



<sup>12</sup>Cynthia Dwork and Guy N Rothblum. "Concentrated differential privacy". In: (2016).

<sup>13</sup>Mark Bun and Thomas Steinke. "Concentrated Differential Privacy: Simplifications, Extensions, and Lower Bounds". In: 2016.

## 34 Lower-bounds and upper-bounds with CDP

**Private projection estimator :**

$$\hat{\pi}^{\text{proj}}(\mathbf{X}) = \sum_{i=1}^N \left( \hat{\theta}_i + C_{\rho, N} \mathcal{N}(0, 1) \right) \phi_i \quad \text{where} \quad \hat{\theta}_i := \frac{1}{n} \sum_{j=1}^n \phi_i(X_j).$$

$\hat{\pi}^{\text{proj}}$  is  $\rho$ -zCDP.

**Resulting upper-bound :**

$$\mathbb{E} \left( \|\hat{\pi}^{\text{proj}}(\mathbf{X}) - \pi\|_{L_2}^2 \right) \leq C_{L, \beta} \max \left\{ n^{-\frac{2\beta}{2\beta+1}}, (n\sqrt{\rho})^{-\frac{2\beta}{\beta+1}} \right\}$$

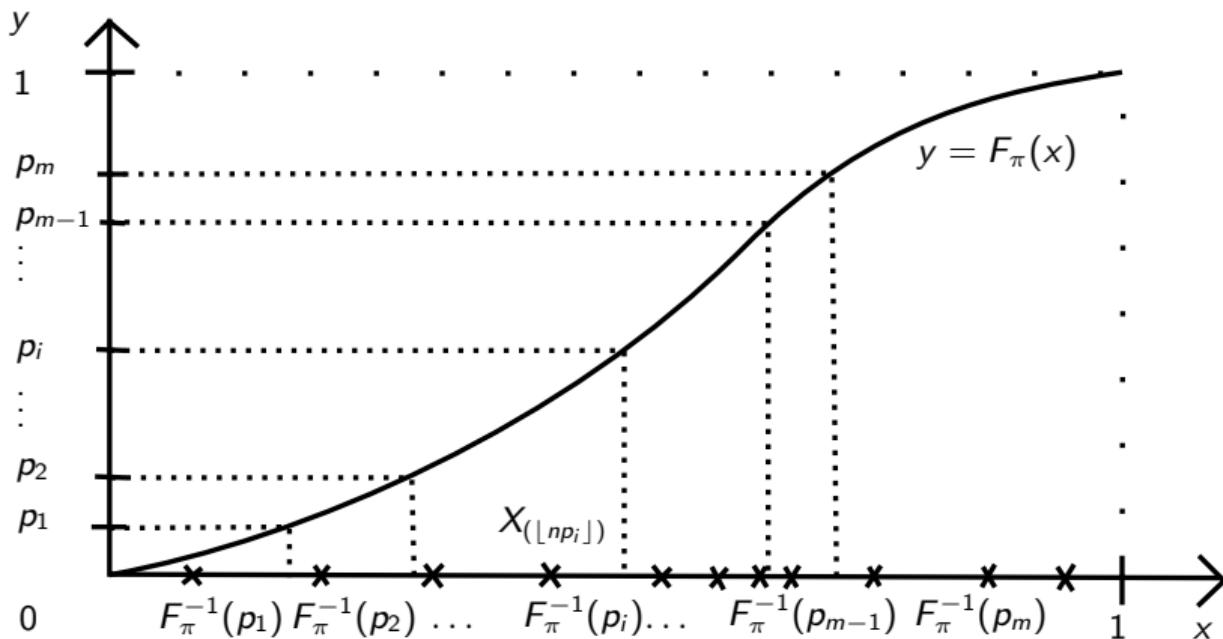
**Lower-bound against  $\rho$ -zCDP estimators :**

$$\inf_A \sup_{\pi} \mathbb{E} \left( \|A(\mathbf{X}) - \pi\|_{L_2}^2 \right) \geq C_{L, \beta} \max \left\{ n^{-\frac{2\beta}{2\beta+1}}, (n\sqrt{\rho})^{-\frac{2\beta}{\beta+1}} \right\}$$

## 35 Quantiles Estimation Problem

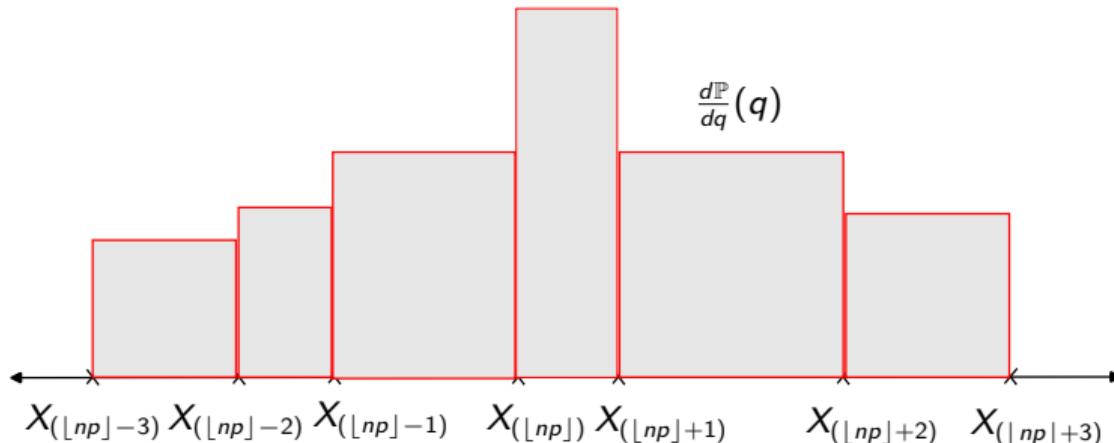
**Inputs** : samples  $\mathbf{X} = (X_1, \dots, X_n) \stackrel{\text{i.i.d.}}{\sim} \mathbb{P}_\pi$   $\mathbf{p} = (p_1, \dots, p_m) \in (0, 1)^m$  sorted.

**Desired output** : Quantile estimator  $\mathbf{q} \in [0, 1]^m$  of  $(F_\pi^{-1}(p_1), \dots, F_\pi^{-1}(p_m))$ .



## 36 Private Exponential Quantiles

**Mechanism :**<sup>14</sup> For a single quantile  $q$  (associated with  $p$ ),



**Concentration result :**<sup>16</sup> When  $\pi$  is away from 0 on a neighborhood of  $F_\pi^{-1}(p)$ ,

$$\mathbb{P}(|q - F_\pi^{-1}(p)| > \gamma) \leq P(n) \max \left( e^{-C_1 \epsilon n \gamma}, e^{-C_2 \gamma^2 n} \right).$$

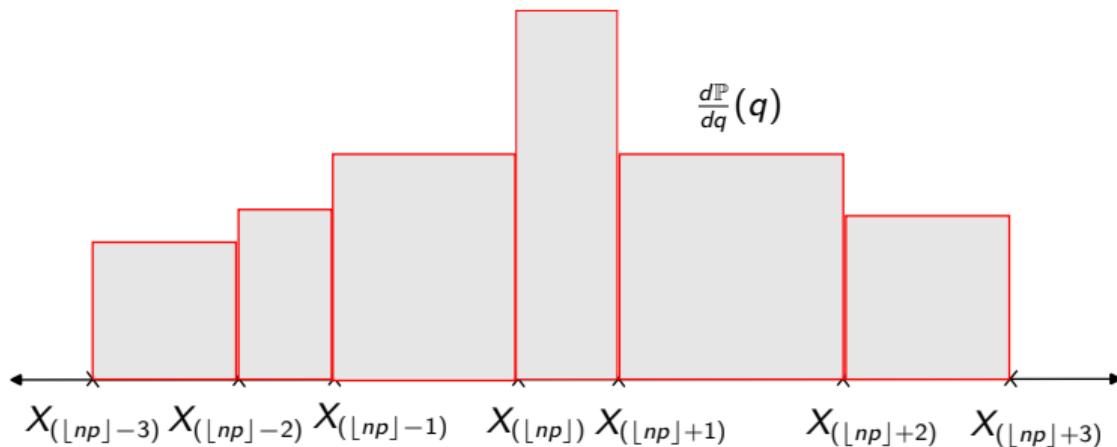
<sup>14</sup> Adam D. Smith. "Privacy-preserving statistical estimation with optimal convergence rates". In: 2011.

<sup>15</sup>  $d\mathbb{P}(q) \propto e^{-\frac{\epsilon}{2} |\{i | X_i < q\}| - \lfloor np \rfloor} dq$

<sup>16</sup> Clément Lalanne, Aurélien Garivier, and Rémi Gribonval. "Private Statistical Estimation of Many Quantiles". In: 2023.

## Independent Private Quantiles

Idea : Use QExp independently on  $\mathbf{p}$  with simple composition.



**Concentration result :**<sup>18</sup> When  $\pi$  is away from 0 on a neighborhood of  $F_\pi^{-1}(\mathbf{p})$ ,

$$\mathbb{P}\left(\|\mathbf{q} - F_\pi^{-1}(\mathbf{p})\|_\infty > \gamma\right) \leq P(n, m) \max\left(e^{-C_1 \frac{\epsilon n \gamma}{m}}, e^{-C_2 \gamma^2 n}\right).$$

---

<sup>17</sup>  $dP(q_i) \propto e^{-\frac{\epsilon}{2m} \left| |\{i | X_i < q_i\}| - \lfloor np_i \rfloor \right|} dq_i$

<sup>18</sup> Lalanne, Garivier, and Gribonval, "Private Statistical Estimation of Many Quantiles".

## 38 Joint Exponential Private Quantiles

**Idea :**<sup>19</sup> Leverage structural dependencies. Quantiles are non-decreasing, between  $q_i$  and  $q_j$  should fall approximately  $n|p_i - p_j|$  points.<sup>20</sup>

**"Fun" discovery :**<sup>21</sup>  $\text{JointExp} \approx \text{Inverse Sensitivity Mechanism}$ <sup>22</sup>.

**Consistency result :** When  $\pi$  is away from 0 on a neighborhood of  $F_\pi^{-1}(\mathbf{p})$ ,  $\text{JointExp}$  is consistent.

---

<sup>19</sup> Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. "Differentially Private Quantiles". In: 2021.

<sup>20</sup>  $d\mathbb{P}(\mathbf{q}) \propto e^{-\frac{\epsilon}{2} \sum_{i=1}^{m+1} |\delta^{\text{JE}}(i, \mathbf{X}, \mathbf{q})|} d\mathbf{q}$  where  $\delta^{\text{JE}}(i, \mathbf{X}, \mathbf{q}) := n(p_i - p_{i-1}) - \#(\mathbf{X} \cap (q_{i-1}, q_i])$

<sup>21</sup> Clément Lalanne et al. "Private quantiles estimation in the presence of atoms". In: (2023). ISSN: 2049-8772.

<sup>22</sup> Hilal Asi and John C. Duchi. "Near Instance-Optimality in Differential Privacy". In: CoRR (2020). arXiv: 2005.10630.

## 39 Recursive Private Quantiles

**Idea :**<sup>23</sup> Use QExp recursively with a dichotomy on  $\mathbf{p}$ .

**Concentration result :**<sup>24</sup> When  $\pi$  is away from 0 on a neighborhood of  $F_\pi^{-1}(\mathbf{p})$ ,

$$\mathbb{P}\left(\|\mathbf{q} - F_\pi^{-1}(\mathbf{p})\|_\infty > \gamma\right) \leq P(n, m) \max\left(e^{-C_1 \frac{\epsilon n \gamma}{(\log_2(m))^2}}, e^{-C_2 \gamma^2 n}\right).$$

**Remark :** Almost polylogarithmic degradation in  $m$  !

---

<sup>23</sup>Haim Kaplan, Shachar Schnapp, and Uri Stemmer. “Differentially Private Approximate Quantiles”. In: ed. by Kamalika Chaudhuri et al. PMLR, 2022.

<sup>24</sup>Lalanne, Garivier, and Gribonval, “Private Statistical Estimation of Many Quantiles”.

## 40 Histograms

Idea :<sup>25</sup>

$$\hat{\pi}^{\text{hist}}(t) := \sum_{b \in \text{bins}} \mathbb{1}_b(t) \frac{1}{nh} \left( \sum_{i=1}^n \mathbb{1}_b(X_i) + \frac{2}{\epsilon} \mathcal{L}_b \right).$$

**Concentration result :**<sup>26</sup> Bins of size  $h$ ,  $\gamma > C_4 h$ ,  $\pi$  is **L-Lipschitz**,  $I$  is a strict sub-interval

$$\begin{aligned} & \mathbb{P}\left(\|F_{\hat{\pi}^{\text{hist}}}^{-1} - F_\pi^{-1}\|_{\infty, I} > \gamma\right) \\ & \leq \frac{1}{h} e^{-C_1 \gamma h n \epsilon} + \frac{2}{h} e^{-C_2 h^2 (C_3 \gamma - Lh)^2 n}. \end{aligned}$$

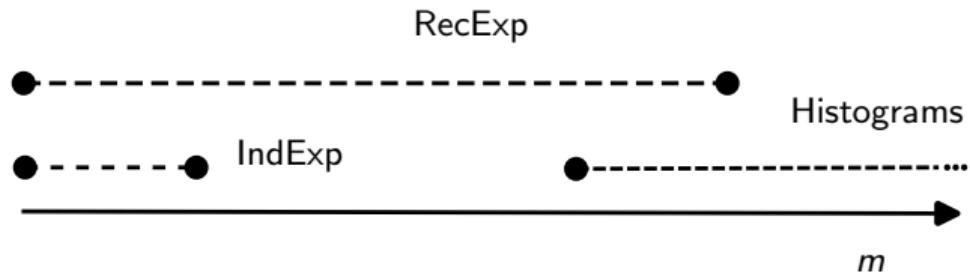
**Remark :** No degradation in  $m$ , but high entry cost.

---

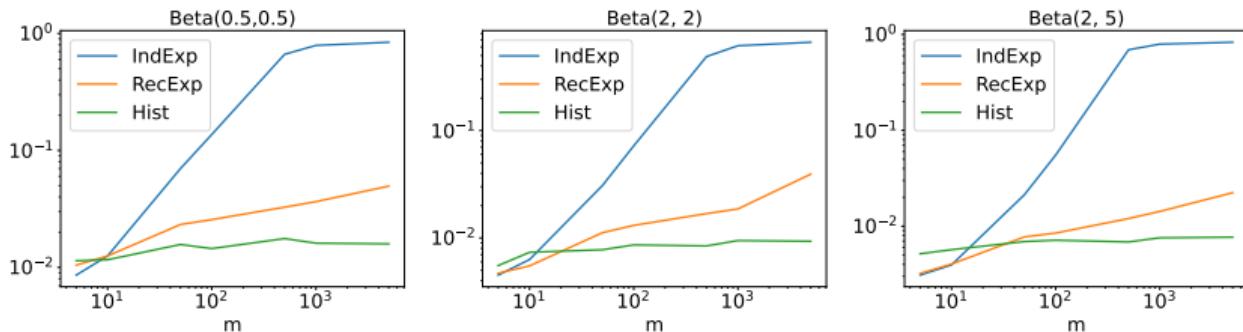
<sup>25</sup>Wasserman and Zhou, "A Statistical Framework for Differential Privacy".

<sup>26</sup>Lalanne, Garivier, and Gribonval, "Private Statistical Estimation of Many Quantiles".

## 41 Theoretical choice of algorithm



## 42 Numerical validation

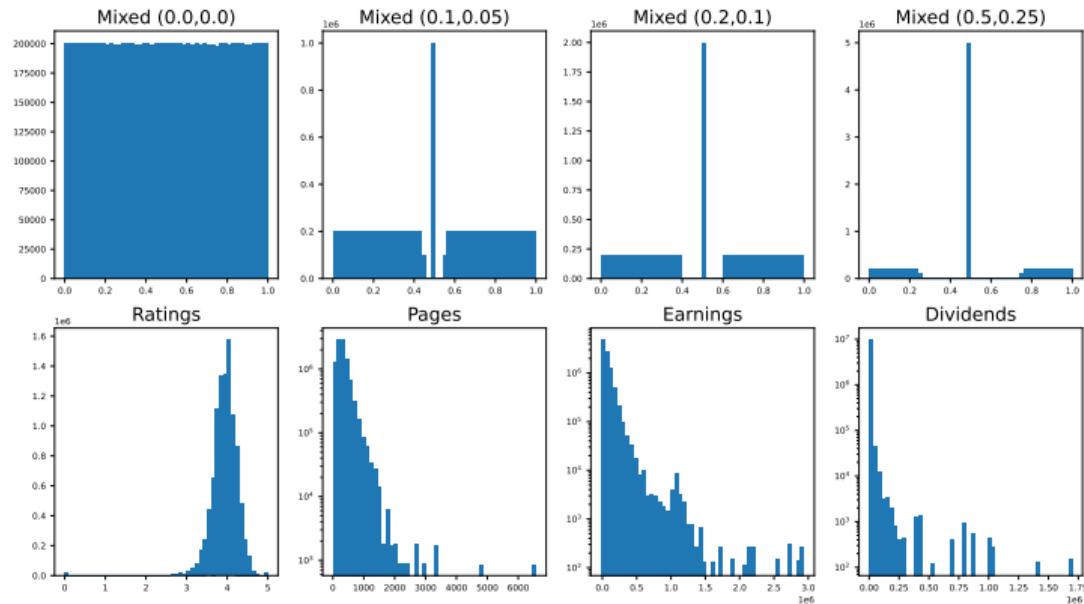


The vertical axis reads the error  $\mathbb{E}(\|\mathbf{q} - F^{-1}(\mathbf{p})\|_\infty)$  where  $\mathbf{p} = \left(\frac{1}{4} + \frac{1}{2(m+1)}, \dots, \frac{1}{4} + \frac{m}{2(m+1)}\right)$  for different values of  $m$ ,  $n = 10000$ ,  $\epsilon = 0.1$ , and  $\mathbb{E}$  is estimated by Monte-Carlo averaging over 50 runs. The histogram is computed on 200 bins.

## 43 Dealing with atomic distributions

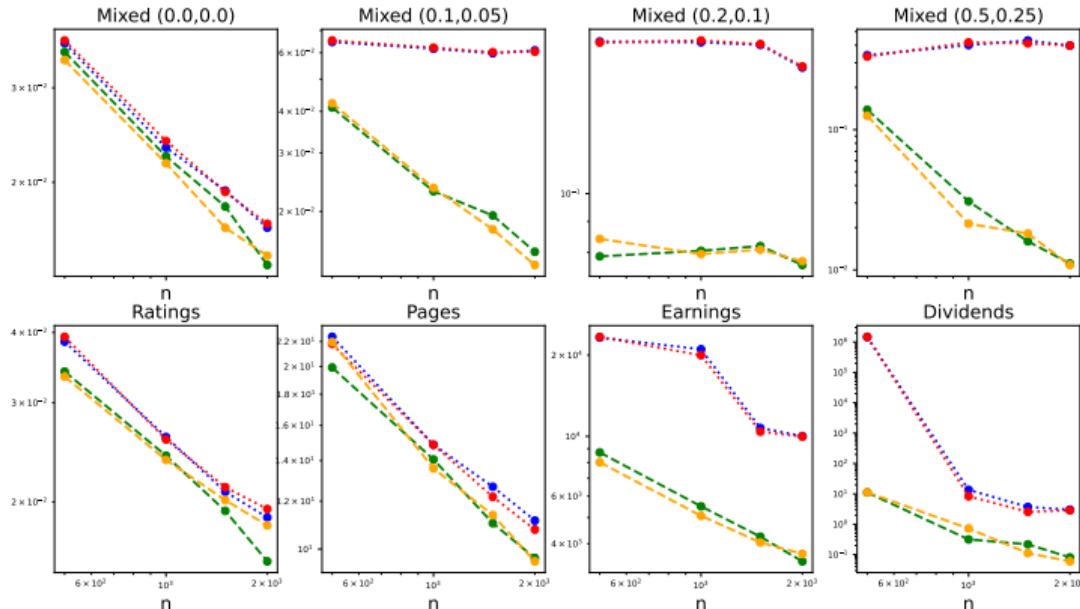
**Inconsistency result :**<sup>27</sup> When dealing with atomic distributions, all the \*Exp mechanisms are inconsistent or have poor performances.

**Proposed solution :** Smoothing the distribution with noise addition can make those mechanisms consistent and helps the performances.



<sup>27</sup>Lalanne et al., "Private quantiles estimation in the presence of atoms".

## 44 Dealing with atomic distributions



The vertical axis reads the error  $\mathbb{E}(\|\hat{\mathbf{q}} - F^{-1}(\mathbf{p})\|_\infty)$  where  $\mathbf{p} = (\frac{1}{m+1}, \dots, \frac{m}{m+1})$  for  $m = 8$ ,  $\epsilon = 1$ ,  $\hat{\mathbf{q}}$  is the private estimator, and  $\mathbb{E}$  is estimated by Monte-Carlo averaging over 50 runs.