The Privacy Funnel from the viewpoint of Local Differential Privacy

Milan Lopuhaä-Zwakenberg
m.a.lopuhaa@tue.nl

Eindhoven University of Technology

International Conference on the Digital Society: Protecting Privacy in Open (& Big) Data Settings

CV page

- future: Postdoc, formal methods & tools, Twente
- **2018–present: Postdoc, security group, Eindhoven**
- 2014–2018: PhD, algebraic geometry, Nijmegen

Interests page

Interests:

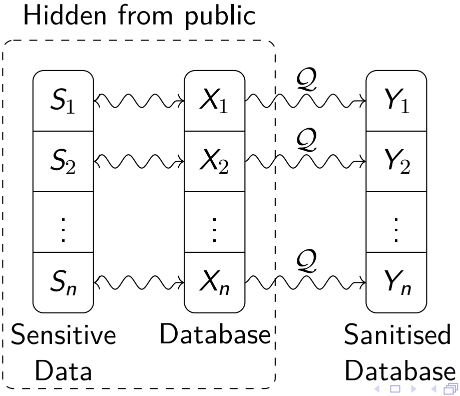
- Data sanitisation
- Differential privacy
- Privacy and utility metrics
- Information-theoretical properties of metrics

Current projects:

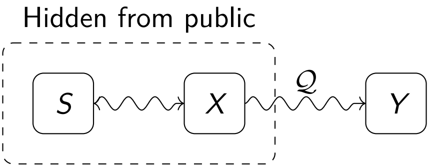
- Robust privacy metrics for Privacy Funnel
- Differentially private classifier learning
- Fisher information in private distribution estimation

Privacy Funnel setting

- Database $\vec{X}$ consisting of rows $X_1, \ldots, X_n$.
- Each entry X_i is correlated to secret information S_i.
- Goal: output sanitised database $\vec{Y}$ that does not leak about $\vec{S}$.
- Method: *probabilistic* protocol $\mathcal{Q}$ operating on rows individually.
- Assumption: (S_i, X_i) discrete random variables, joint distr. known.



Privacy Funnel setting



- Goal: Y contains lot of info about X, little info about S. *How do we measure these?*
- Typically: Find $\mathcal{Q}$ that maximises $\mathrm{I}(X;Y)$ while $\mathrm{I}(S;Y) \leq L$ for a given L.
- Problems:
  - $\mathrm{I}(S;Y)$ is *average* leakage, some rows may be more compromised
  - Current methods only give local optima

Local Differential Privacy & Local Information Privacy

Definition

Let $\varepsilon > 0$.

1. $\mathcal{Q}$ satisfies ε-LDP (w.r.t. S) if for all $s, s' \in \mathcal{S}$ and all $y \in \mathcal{Y}$:

$$\frac{\mathbb{P}(\mathcal{Q}(X) = y | S = s)}{\mathbb{P}(\mathcal{Q}(X) = y | S = s')} \leq \mathrm{e}^{\varepsilon}$$

2. $\mathcal{Q}$ satisfies ε-LIP (w.r.t. S) if for all $s \in \mathcal{S}$ and all $y \in \mathcal{Y}$:

$$\mathrm{e}^{-\varepsilon} \leq \frac{\mathbb{P}(\mathcal{Q}(X) = y | S = s)}{\mathbb{P}(\mathcal{Q}(X) = y)} \leq \mathrm{e}^{\varepsilon}$$

Privacy depends on ε; typically $\varepsilon \approx 1$.

Properties:

- Worst-case metrics
- LDP stricter than LIP
- both depend on $\mathrm{p}_{S,X}$
- $\frac{\mathrm{p}_{y|s}}{\mathrm{p}_y} = \frac{\mathrm{p}_{s|y}}{\mathrm{p}_s}$, so LIP bounds difference between prior and posterior
- LDP does so even when p_S is unknown
- *but we assume it is known*

Definition

ε-LDP: $\forall y, s, s'$:

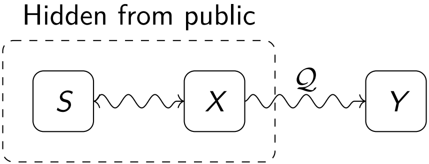
$$\frac{\mathrm{p}_{y|s}}{\mathrm{p}_{y|s'}} \leq \mathrm{e}^{\varepsilon}$$

ε-LIP: $\forall y, s$:

$$\mathrm{e}^{-\varepsilon} \leq \frac{\mathrm{p}_{y|s}}{\mathrm{p}_y} \leq \mathrm{e}^{\varepsilon}$$

Conclusion: LIP more sensible metric than LDP

Problems



Problem

Given ε, find $\mathcal{Q}$ that maximises $\mathrm{I}(X;Y)$ s.t. $\mathcal{Q}$ satisfies ε-LDP/LIP.

I will show:

- We can find the optimal $\mathcal{Q}$;
- This is (computationally) easier for LIP than for LDP.

Optimal ε-LDP protocol

Let $a = |\mathcal{X}|$, $b = |\mathcal{Y}|$.

- $\mathcal{Q}$ is given by $G := \mathrm{p}_{Y|X} \in \mathbb{R}^{b \times a}$.
- Theorem: $b = a$ for optimal protocol.
- $\{\varepsilon\text{-LDP protocols}\} = \Delta$ where

$$\Delta = \left\{ G \in \mathbb{R}^{b \times a} : \begin{array}{c} \forall x: \sum_y G_{y|x} = 1, \\ \forall x,y: G_{y|x} \geq 0, \\ \forall s \neq s', y: \sum_x G_{y|x} \mathrm{p}_{x|s} \leq \mathrm{e}^{\varepsilon} \sum_x G_{y|x} \mathrm{p}_{x|s'} \end{array} \right\}$$

  This is a polyhedron of dimension $a^2 - a$.
- To do: maximise $\mathrm{I}(X; Y) = \sum_{x,y} \mathrm{p}_x G_{y|x} \log \frac{G_{y|x}}{\sum_{x'} \mathrm{p}_{x'} G_{y|x'}} =: f(G)$ over Δ.
- f is convex, so maximum is obtained in vertex.
- **Find vertices of Δ $\Rightarrow$ find optimal ε-LDP protocol**

Optimal ε-LIP protocol

Let $a = |\mathcal{X}|$, $b = |\mathcal{Y}|$.

- $\mathcal{Q}$ is given by $J := \mathrm{p}_{X|Y} \in \mathbb{R}^{a \times b}$ and $\vec{\theta} = \mathrm{p}_Y \in \mathbb{R}^b$ with $J \cdot \vec{\theta} = \mathrm{p}_X$.
- $\forall y : J_y \in \Gamma$ where

$$\Gamma = \left\{ R \in \mathbb{R}^a : \substack{\sum_x R_x = 1, \\ \forall x: R_x \geq 0, \\ \forall s: \mathrm{e}^{-\varepsilon}\, \mathrm{p}_s \leq \sum_x \mathrm{p}_{s|x} R_x \leq \mathrm{e}^{\varepsilon}\, \mathrm{p}_s} \right\}$$

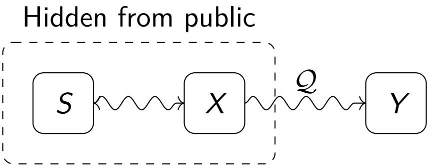
  This is a polyhedron of dimension $a - 1$.
- Theorem: optimal $\mathcal{Q}$ has J_y vertex of Γ for all y.
- Theorem: if we know vertices of Γ we can find $\vec{\theta}$ via linear optimisation problem.
- **Find vertices of $\Gamma \Rightarrow$ find optimal ε-LIP protocol**

LDP vs LIP

Let $a = |\mathcal{X}|$, $c = |\mathcal{S}|$

- Complexity vertex enumeration: $\mathcal{O}(ndv)$, with $(n, d, v) = (\text{dimension}, \text{inequalities}, \text{vertices})$.
- LDP: $n = a^2 - a$, $d = a^2 + c^2 - c$
- LIP: $n = a - 1$, $d = a + 2c$
- v unknown, generally $v \leq \binom{n}{d}$
- So: LIP faster! (about 5000× for $c = 2, a = 5$)
- Both methods computationally infeasible for large a, c

Conclusion



We have solved:

Problem

Given ε, find $\mathcal{Q}$ that maximises $\mathrm{I}(X;Y)$ s.t. $\mathcal{Q}$ satisfies ε-LDP/LIP.

- Better privacy guarantees than original Privacy Funnel, with optimal utility
- LIP faster to optimise, more sensible than LDP
- Optimisation hard for large spaces