V) Rethinking Disclosure Prevention with PML

1) Differential Privacy

How to keep anonymity when computing statistics about medical records, personal incomes, survey responses, ... ?

Let $\mathcal{X}$ be the set of values that individuals can have, and let $\mathcal{D} = \mathcal{X}^N$ where N is the number of individuals.
d and d' of $\mathcal{D}$ are called neighbors if they differ in only one individual.

Let $\mathcal{Y}$ be the set of possibly outputted values and let Y be the outputted value characterized by $\mathbb{P}_{Y|D}$. Since $\mathcal{Y}$ can be infinite, we call $\mathcal{T}$ the set of measurable sets by any $\mathbb{P}_{Y|D=d}$. We call $\mathcal{M}(d)$ the random variable $Y|D=d$. We say that $\mathcal{M}$ is ε-differentially private iff $\forall S \in \mathcal{T}, \log\left(\frac{\mathbb{P}(\mathcal{M}(d) \in S)}{\mathbb{P}(\mathcal{M}(d') \in S)}\right) \leq \varepsilon$, for any pair of neighbors (d, d').

k-anonymity characterizes a dataset $d \in \mathcal{D}$ and it is just the fact that $\forall x \in \mathcal{X}, |\{i \in [N] \mid d_i = x\}|$ is either bigger than k or equal to zero.

↳ This way we can release data about what do people vote or what is the frequency of a given disease, right ? → Well we need to be cautious; what if k persons voted for the same person and live in the same place (like a family) then we can identify them. We introduce l-diversity : we add the condition that the groups should have diversity so we can't identify them all.

We can represent a data-set by a table :

- id : Not directly in the matrix, this is only the number of the line
- Quasi-Identifiers : Zip code, Age, ...
- Sensitive-Attributes : Disease, Vote, ...

id	QI	SA

T =

k-anonimity only is for QI, so if there is a leak, we cannot know what are your SA, unless all k have the same SA.

We can easily define the k-anonimity equivalence classes by $a \sim b$ iff $QI[a]=QI[b]$.

k-anonimity $\Leftrightarrow \forall EC, |EC| \geq k$ (EC means Equivalence Class)

l-diversity $\Leftrightarrow \forall EC, |SA(EC)| \geq l$ (implies l-anonimity)

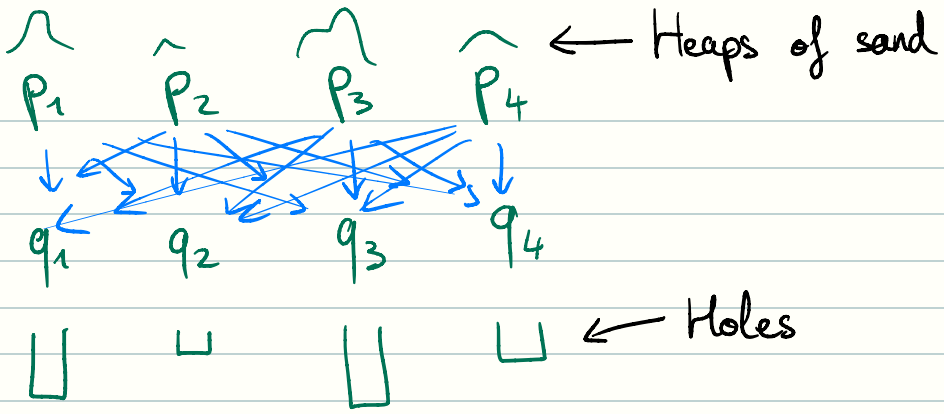
↳ One can find other variants of l-diversity.

l-diversity is still weak : imagine an EC with 99% of a very rare SA, then it leaks very strong information. The empirical distribution of SA(EC) must be close to the whole database SA distribution.

The data-base T satisfies t-closeness according to some distance $D : \Delta(SA)^2 \to \mathbb{R}_+$ iff $\forall EC, D(\mathbb{P}^{(emp)}_{SA|i \in EC}, \mathbb{P}^{(emp)}_{SA}) \leq t$. Emp means "empirical".

<u>Remark:</u> t-closeness gives l-diversity and k-anonimity guarantees (for every distance?).

<u>Example:</u> Given $P=(p_1,\dots,p_k)$ and $Q=(q_1,\dots,q_k)$ element of $\Delta(SA)$, a transport plan is a $k \times k$ matrix γ such that γ_{ij} shows the amount that we take from p_i to put it on q_j (could be negative but this is useless). This transport plan is said to be valid if $\begin{cases} \forall i \in [k], \sum_j \gamma_{ij} = p_i \\ \forall j \in [k], \sum_i \gamma_{ij} = q_j \end{cases}$.



The set of valid transport matrices is called $\Gamma(P \to Q)$. For instance, we can simply check that $\Gamma(P \to Q) = \Gamma(Q \to P)^T$.

Finally, we define a function between the heaps $(x_i)_{i \in [k]}$ and the holes $(y_j)_{j \in [k]}$. The Earth Mover's Distance or the Optimal Transport Distance, is defined by:

$$EMD(P, Q) = \inf_{\gamma \in \Gamma(P \to Q)} \sum_{i,j} \gamma_{ij} d(x_i, y_j).$$

A particular case of this, is if we only have a distance $d : SA^2 \to \mathbb{R}_+$ on SA, then

$$EMD(P, Q) = \inf_{\gamma \in \Gamma(P \to Q)} \sum_{i,j} \gamma_{ij} d(i, j).$$

2) Preliminaries

Again we recall the setup:

useful secret output guess

$U - X - Y - \hat{U}$

$P_{U|X}$ P_X $P_{Y|X}$ $P_{\hat{U}|Y}$ → Imagine those as deterministic.

The min-entropy is defined as $H_\infty(X) = -\log(\max_x P_X(x))$ or, accordingly to its name, $H_\infty(X) = \log(\min_x \frac{1}{P_X(x)})$.

The min-leakage is given by $L_\infty(X, Y) = H_\infty(Y) - H_\infty(Y|X)$.

The min-capacity is given by $C_\infty(\mathbb{P}_{Y|X}) = \sup_{\mathbb{P}_X} L_\infty(X,Y)$.

Let $(\Omega, \mathcal{T}, \mu)$ be a measure space and let P and Q random variables $\ll \mu$ of Ω with respective pmf p, q. The Renyi Divergence of order ∞ of P from Q is defined by $D_\infty(P \| Q) = \operatorname{ess\,sup}\left[\log(p) - \log(q)\right]$ assuming that $(-\infty) - (-\infty) = 0$.

We recall that Maximal Leakage is given by:

$$\mathcal{L}(X \to Y) = \inf_{\mathbb{P}_Y^*} D_\infty(\mathbb{P}_X \mathbb{P}_Y^* \| \mathbb{P}_{X,Y})$$

↳ Surprisingly, for a fixed support of X, this only depends on $\mathbb{P}_{Y|X}$

Well Pointwise Maximal Leakage is given by:

$$\ell(X \to y) = D_\infty(\mathbb{P}_{X|Y=y} \| \mathbb{P}_X)$$

Let $\mathcal{P}$ be a subset of all possible distributions for X. A channel $X \to Y$ is said to satisfy $(\varepsilon, \mathcal{P})$-PML if $\forall \mathbb{P}_X \in \mathcal{P}$, $\mathbb{P}_Y(\{y \in \mathcal{Y} \mid \ell(X \to y) \leq \varepsilon\}) = 1$

or, equivalently, $\sup_{\mathbb{P}_X \in \mathcal{P}} D_\infty(\mathbb{P}_{XY} \| \mathbb{P}_X \mathbb{P}_Y) \leq \varepsilon$.

Proof of the equivalence: $D_\infty(\mathbb{P}_{XY} \| \mathbb{P}_X \mathbb{P}_Y) = \operatorname{esssup} \log i(x;y)$ and $\ell(X \to y) = \max_x i(x,y)$, so we just follow the definitions.

If the density of Y is continuous for a given $X = x$, then the two conditions above also are equivalent to $\forall \mathbb{P}_X \in \mathcal{P}$, $\forall y \in \mathcal{Y}$, $\ell(X \to y) \leq \varepsilon$.

A channel $X \to Y$ is said to satisfy ε-PML if $\max_{x \in \mathcal{X}} i(x,y) \leq \varepsilon$ almost everywhere on $\mathcal{Y}$, i.e. $D_\infty(\mathbb{P}_{XY} \| \mathbb{P}_X \mathbb{P}_Y) \leq \varepsilon$, for the real distribution of x.

The leakage capacity can be defined as

$$C_\mathcal{L}(X \to Y) = \max_{x,x'} \sup_y \log \frac{\mathbb{P}_{Y|X=x}(y)}{\mathbb{P}_{Y|X=x'}(y)}$$

Why sup instead of essup ?

Can be infinite

Theorem: $\sup_{\mathbb{P}_X} D_\infty(\mathbb{P}_{XY} \| \mathbb{P}_X \mathbb{P}_Y) = C_\mathcal{L}(X \to Y)$

Proof: On the LHS, we have $\log \frac{\mathbb{P}_{Y|X=x}(y)}{\mathbb{P}_Y(y)}$ (1). On the RHS, we have $\log \frac{\mathbb{P}_{Y|X=x}(y)}{\mathbb{P}_Y(y)} - \log \frac{\mathbb{P}_{Y|X=x'}(y)}{\mathbb{P}_Y(y)}$ (2). LHS ≤ RHS is clear since we can easily find an x' such that $\mathbb{P}_{Y|X}(y|x') \leq \mathbb{P}_Y(y)$. We now want LHS ≥ RHS.

Let $y^* \in \mathcal{Y}$ as close to the supremum as we need, and let x_{min} and x_{max} the values maximizing the expression. Then we chose the following $\mathbb{P}_X$: $\begin{cases} \mathbb{P}_X(x_{min}) = 1-\alpha \\ \mathbb{P}_X(x_{max}) = \alpha \end{cases}$

Then LHS ≥ $\max_x \log \frac{\mathbb{P}_{Y|X}(y^*|x)}{\mathbb{P}_Y(y^*)} = \log \frac{\mathbb{P}_{Y|X}(y^*|x_{max})}{\alpha \mathbb{P}_{Y|X}(y^*|x_{max}) + (1-\alpha)\mathbb{P}_{Y|X}(y^*|x_{min})}$ → Definition of x_{max}

The limit when $\alpha \to 0$ is RHS □

(1) Depends on $\mathbb{P}_X$ (2) Does not depend on $\mathbb{P}_X$

3) Impossibility of Absolute Disclosure Prevention

Setup of Dwork and Naor (2010): - $\mathbb{P}_X$ is known

- The classic Markov chain $U - X - Y$:
  - $\mathbb{P}_{U|X}$ is known
  - $H(\mathbb{P}_U)$ is large
  - $\exists y$, $H(\mathbb{P}_{U|Y=y})$ is small
- There is another Markov chain $W - X - Y$:
  - $H(W) < H(U)$
  - $H(\mathbb{P}_W)$ is large
  - $\forall y$, $H(\mathbb{P}_{W|Y=y})$ is large

Global = Small entropy i.e. less useful information that is commonly known.

Local = High entropy i.e. very rare and useful information.

Our setup: Prior belief $Q_X \in \Delta(\mathcal{X})$. But actually, we don't care. → Worst case: $Q = \mathbb{P}$.

We say that the channel $X \to Y$ discloses the value of U iff $\inf_y H_\infty(\mathbb{P}_{U|Y=y}) = 0$.

Theorem 3.5: $\varepsilon \geq 0$

Assume $H_\infty(P_U) > \varepsilon$. If $D_\infty(\mathbb{P}_{XY} \| \mathbb{P}_X \mathbb{P}_Y) \leq \varepsilon$, then any disclosure is impossible for any prior.

Remarks: i) It does not mean that we leak the same information to all priors. If $l_Q(X \to y) > l(X \to y)$ then we give more info to Q. If $l(X \to y) > l_Q(X \to y)$ then that's a pity since Q will not realize how rare is this information

ii) We only care of the true prior $\mathbb{P}$.

iii) The condition of the theorem simply is $(\varepsilon, \{\mathbb{P}_X\})$-PML. However, if $\mathbb{P}_X$ is unknown but we know that $\mathbb{P}_X$ is in some $\mathcal{P}$, then the theorem remains valid with $(\varepsilon, \mathcal{P})$-PML guaranty.

Theorem 3.7 (Absolute Disclosure Prevention): Let V be a non-constant but deterministic function of X. Therefore,

$$H_\infty(\mathbb{P}_{V|Y=y}) \geq \log\left(1 + \frac{\min_x q_X(x)}{1 - \min_x q_X(x)} e^{-C_\ell(X \to Y)}\right)$$

So if $C_\ell(X \to Y) < +\infty$ then any disclosure is impossible.

Note: $\mathbb{P}_{V|Y=y}$ can be replaced by any adversarial prior $Q_{V|Y=y}$.

Understanding theorem 3.5: U is what we call a feature of X. The main condition is $(\varepsilon, \mathcal{P})$-PML with $\mathbb{P}_X \in \mathcal{P}$. It splits the features of X in two categories:

$H_\infty(\mathbb{P}_U) > \varepsilon$	$H_\infty(\mathbb{P}_U) \leqslant \varepsilon$
Local features	Global features
"Cannot be disclosed"	"May be disclosed"

How to pick ε? Let x_{min} be the least likely realization of X, $p_{min} = \mathbb{P}_X(x_{min})$. Let $\varepsilon^* = \log \frac{1}{1 - p_{min}}$. Assume $\mathbb{P}_{Y|X}$ satisfies $(\varepsilon, \{\mathbb{P}_x\})$-PML.

- If $\varepsilon < \varepsilon^*$, then no feature is disclosable ; $C_\chi(X \to Y) < +\infty$.
- If $\varepsilon \geqslant \varepsilon^*$, then there exist a disclosable feature of X.

Proof: i) Assume $C_\chi(X \to Y) = +\infty$. We recall that:

$$C_\chi(X \to Y) = \max_{x, x'} \sup_y \log \frac{\mathbb{P}_{Y|X=x}(y)}{\mathbb{P}_{Y|X=x'}(y)}.$$

Therefore, there exist $y_\varepsilon \in \mathcal{Y}$ such that

$$\max_{x, x'} \frac{\mathbb{P}_{Y|X=x}(y)}{\mathbb{P}_{Y|X=x'}(y)} \geqslant \frac{1}{\varepsilon}$$

Then, we can easily show that $\operatorname*{ess\,sup}_y \ell(X \to y) \geqslant \varepsilon^*$.

ii) Let $U = \delta_X^{x_{min}}$, and assume $\varepsilon = \varepsilon^*$. Let $\alpha > 0$. We construct the following channel:
- if $X = x_{min}$, then $Y = 1$
- if $X \neq x_{min}$, then $Y \hookrightarrow \mathcal{B}(1-\alpha)$

The more α is small, the less $\mathbb{P}_{Y|X}$ leaks about x_{min}. For α small enough, we have $(\varepsilon, \mathbb{P}_X)$-PML, and we easily verify that

$P_{Y|U}(0|u) = \begin{cases} 0 & \text{if } u=0 \\ \alpha & \text{if } u=1 \end{cases}$. So if we observe $Y=1$, we reduced U to a single point. $H(U|Y=1) = 0$ and we're done. □

We say that $X \to Y$ singles out if it discloses the value of X:

$$\inf_y H_\infty(P_{X|Y=y}) = 0$$

If we have (ε, P)-PML then if $\varepsilon < H_\infty(X)$, we are guaranteed the channel does not single out.

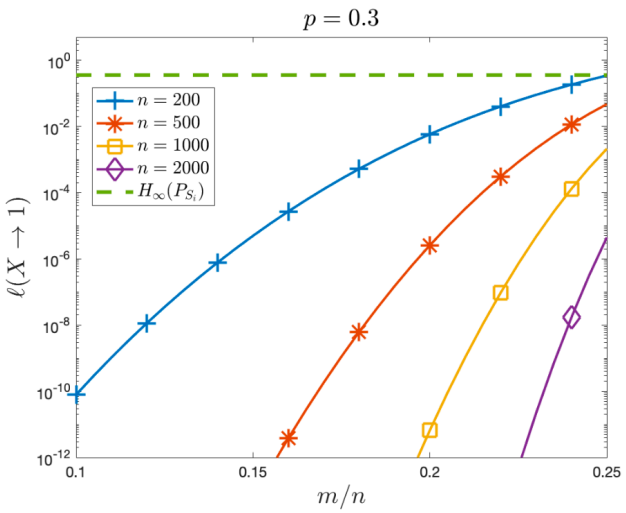
Example 3.11: $X = (D_1, \dots, D_n)$ Let $S_i = 1$ iff i is a women, and $Y =$ "More than m women?" then $S_i - X - Y$.

We assume that $p = 0.3$ and $\frac{m}{n} \leq p$. We have that

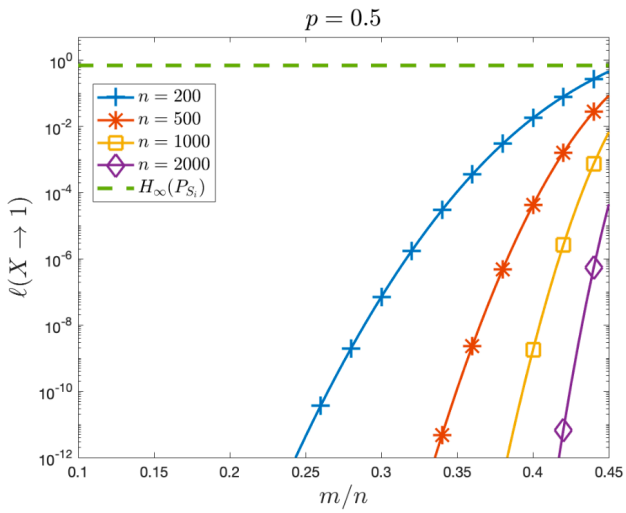
$$\ell(X \to 1) = \log \max_x \frac{P_{Y|X=x}(1)}{P_Y(1)} = \log \frac{1}{P_X(\text{more than } m \text{ women})} = -\log\left(1 - \sum_{k=0}^{m} p^k (1-p)^{n-k} \binom{k}{n}\right)$$

Let $i \in [n]$. $H_\infty(P_{S_i}) = \log\left(\frac{1}{1-p}\right) \approx 0.36$.

The gender of no individual will be disclosed as long as $\begin{cases} \ell(X \to 1) \leq 0.36 \\ \ell(X \to 0) \leq 0.36 \end{cases}$.



(A) Leakage bounds when $p = 0.3$.



(B) Leakage bounds when $p = 0.5$.