Lecture 25
4-21-26

Today

- Review: inverse sensitivity
- Privacy Wrapper Theorem
- Monotone Functions & Interior Points

Inverse Sensitivity mechanism

Goal: approximate function $f: \mathcal{X}^n \to \mathcal{Y}$, $|\mathcal{Y}| < \infty$

$$p(y) \propto \exp\left(-\frac{\varepsilon}{2} \text{len}_f(y; x)\right)$$

$$\text{len}_f(y; x) = \min_{x'} \{ d_H(x, x') \mid f(x') = y \}$$

Claim: This is ε-DP.

Proof Fix y, $x \sim x'$. Then

$$\begin{aligned} \text{len}_f(y; x) &= \min_{x''} \{ d_H(x, x'') \mid f(x'') = y \} \\ &\le \min_{x''} \{ d_H(x', x'') + d(x, x') \mid f(x'') = y \} \\ &\le 1 + \text{len}_f(y; x') \end{aligned}$$

Sensitivity bounded by one $\Rightarrow$ exponential mechanism is ε-DP $\square$

Canonical application: median

Accuracy guarantee:

Def The k-neighborhood of x is

$$\mathcal{N}_k(x) = \{x' : d_H(x, x') \le k\}$$

Theorem [Asi & Duchi 2020] For $k = \frac{2}{\varepsilon}\log\left(\frac{2|\mathcal{X}|^2}{\beta\varepsilon}\right)$, wp $1-\beta$ the inverse sensitivity mechanism returns $\tilde{y}$ such that

$$\min_{x' \in \mathcal{N}_k(x)} f(x') \le \tilde{y} \le \max_{x' \in \mathcal{N}_k(x)} f(x)$$

Very simple algorithm!

Optimal in a strong sense, but major drawbacks:

- unclear how to compute $\text{len}_f(y; x)$ when $|\mathcal{X}| = \infty$
- for many tasks, asymmetry between removal & addition; when you

can add points, can add outliers.

Privacy Wrappers

Def The *k-down neighborhood* of x is

$$\mathcal{N}_k^{\downarrow}(x) = \{x' : x' \subseteq x,\ |x'| \geq |x| - k\}$$

Theorem [Linder, Raskhodnikova, Smith, Steinke 2025]
There is an (ε, δ)-DP mechanism that, for any $f: \mathcal{X}^n \to \mathcal{Y}$, for

$k = \frac{\ln 1/\delta}{\varepsilon} \cdot 2^{O(\log^* |\mathcal{Y}|)}$, with probability 1 returns $\tilde{y}$ such that

$$\min_{x' \in \mathcal{N}_k^{\downarrow}(x)} f(x) \leq \tilde{y} \leq \max_{x' \in \mathcal{N}_k^{\downarrow}(x)} f(x')$$

Moreover, this algorithm only queries f on datasets in $\mathcal{N}_k^{\downarrow}(x)$; running time

$n^{O(\frac{1}{\varepsilon}\log 1/\delta)} \cdot$ (time to compute f)

Remarks:

- Algorithm more complicated than inverse sensitivity
- Fixes drawbacks: always computable with query access to f, always accuracy no worse.
- LRSS25 also give pure-DP version which succeeds wp $1-\beta$ and has

$$k = O\left(\frac{1}{\varepsilon}\log(|\mathcal{X}|/\beta)\right)$$

- Looks familiar to algorithmic robust statistics: "for all large subset..."

Open Directions

- What is the analog of this for multivariate outputs?
- When can this be implemented efficiently?
- when can we accommodate $|\mathcal{Y}| = \infty$?

Analysis: Monotonicity & Shifted Inverse

Fang, Dong, Yi 2022 give an algorithm for monotone functions

Def A function $f: \mathcal{X}^n \to \mathcal{Y}$ is monotone if $x \leq x'$ implies $f(x) \leq f(x')$.

Theorem [FDY 22] For any monotone $f: \mathcal{X}^n \to \mathcal{Y}$, the Shifted Inverse (Sensitivity) Mechanism is ε-DP and for $k = O\left(\frac{1}{\varepsilon} \log \frac{|\mathcal{Y}|}{\beta}\right)$ w.p. $1-\beta$ returns $\tilde{y}$ such that

$$\min_{x' \in \mathcal{N}_k^{\downarrow}(x)} f(x) \leq \tilde{y} \leq f(x)$$

Shifted inverse defines loss

$$\ell_f(y; x) = \min\{d_H(x, x') : x' \leq x, f(x') \geq y\}$$

Claim For monotone f, any $y \in \mathcal{Y}$, and $x \sim x'$,

$$|\ell_f(y; x) - \ell_f(y; x')| \leq 1$$

Proof: elegant, go look it up! □

Run exponential mechanism to find $\tilde{y}$ approximately minimizing

$$|\ell_f(\tilde{y}; x) - k|$$

For the main result in LRSSZ5, take this algorithm and

1) median $\Rightarrow$ "generalized interior point"

2) Take any f and "monotonize" it.

Informally:

$$M_\ell[f](x) = \max\{f(s) : s \le x,\ |s| \ge |x| - \ell\}$$

More detail on inverse sensitivities

The exponential mechanism we've seen for the median uses loss

$$\ell_{\text{median}}(y; x) = \left|\frac{n}{2} - \#\{i : x_i \le y\}\right|$$

This is exactly "the number of points we need to change in x to make y a median point." Thus, we can see the connection to inverse sensitivity.

Here is another view on the median's loss function: it's a shifted version of the maximum. We have

$$\ell_{\max}(y;x) = \#\{i : x_i \le y\}$$

as a measure of how well y approximates the maximum of x. We measure the approximation by counting data points.

Note that the maximum is a monotone function. We can regard the median* as a shifted version of the maximum.

*or any quantile

FDY use loss function

$$\ell_f(y;x) = \min\{d(x,x') : x' \le x,\ f(x') \le y\}$$

and run the exponential mechanism to minimize $|\ell_f(y;x) - k|$ for

some parameter k.

We've already seen this! Our loss function for the max is

$$\ell_{max}(y; x) = \#\{i : x_i \le y\}$$

$$= \min\{d(x, s) : s \le x,\ \max(s) \le y\}$$

And we've seen how to shift it.

References

Asi & Duchi 2020 https://arxiv.org/abs/2005.10630

Fang, Dong, & Yi 2022 https://juanru-fang.github.io/ShiftedInverse.pdf

Steinke 2023 blog post https://differentialprivacy.org/down-sensitivity/

LRSS 2025 https://arxiv.org/abs/2503.19268