Beyond global sensitivity:

Def: The local sensitivity of a query $f: \mathcal{D} \to \mathbb{R}$ at database D is

$$\Delta_{LS}(f, D) = \max_{D' \sim D} |f(D) - f(D')|$$

- Would like to claim can just replace global sensitivity w/ LS.
- Unfortunately can't do this: magnitude of noise might leak info!

Ex: f = median

$D = (\underbrace{0, 0, 0, \ldots, 0}_{\frac{n}{2}+1}, 10^6, 10^6, \ldots, 10^6)$

$D' = (\underbrace{0, 0, \ldots, 0}_{\frac{n}{2}}, 1^6, 10^6, \ldots)$

$f(D) = f(D') = 0$

$\Delta_{LS}(f, D) = 0 \qquad \Delta_{LS}(f, D') = 10^6$

So adding $Lap\left(\frac{\Delta_{LS}(f,D)}{\varepsilon}\right)$ noise definitely not DP:

on D always return 0, on D' returns large values.

Can we still save this idea?

Basic idea: computing D_{CS} looks at database! Need to privatize.

Propose-Test-Release:

Idea: - Propose bound on LS

- Test if LS below bound (privately)
- If yes, answer w/ (C) noise
  no, answer $\perp$ (or use global sensitivity noise).

Formally:

1) Propose bound β on LS
2) Compute γ = distance from D to D' s.t. $\Delta_{CS}(f, D') \geq \beta$
   (where distance = # points that change). $\Delta\gamma = 1$!
3) Compute $\hat{\gamma} = \gamma + \underbrace{Lap(1/\varepsilon)}_{Z}$
4) If $\hat{\gamma} \leq \frac{1}{\varepsilon}\ln(1/\delta)$ return $\perp$
5) else ($\hat{\gamma} > \frac{1}{\varepsilon}\ln(1/\delta)$) return $f(D) + Lap(\beta/\varepsilon)$

Note: Step 2 can be computationally difficult! But often easy, e.g., median

Thm: PTR is $(2\varepsilon, \delta)$-DP.

Intuition: - $\Delta\gamma = 1$, so $\hat{\gamma}$ DP by Laplace Mechanism
- Step 5 uses bound on LS that is good for both D, D'

Pf: Let $D \sim D' \in \mathcal{D}$

$\Delta\hat{f} = 1$, so $\Pr[M(D) = \perp] \leq e^{\varepsilon} \Pr[M(D') = \perp]$

Two cases:

Case 1: $\Delta_{CS}(f, D) > \beta$

$\Rightarrow \gamma = 0$

Let $S \subseteq \mathbb{R} \cup \{\perp\}$

$\Pr[M(D) \in S] = \Pr[M(D) \in S \cap \{\perp\}] + \Pr[M(D) \in S \cap \mathbb{R}]$

$\leq e^{\varepsilon} \Pr[M(D') \in S \cap \{\perp\}] + \Pr[M(D) \in \mathbb{R}]$

$\leq e^{\varepsilon} \Pr[M(D') \in S] + \Pr_{Z \sim Lap(1/\varepsilon)}[Z \geq \frac{1}{\varepsilon} \ln \frac{1}{\delta}]$

$\leq e^{\varepsilon} \Pr[M(D') \in S] + \delta$ ✓

Case 2: $\Delta_{CS}(f, D) \leq \beta$

$\Rightarrow$ basic composition of 2 mechanisms:

- Laplace def $\hat{f}$: $(\varepsilon, 0)$-DP
- Laplace to release $f(D) + Lap(\beta/\varepsilon)$

$\Rightarrow \beta \geq \Delta_{CS}(f, D)$ by case def

$\Rightarrow \varepsilon$-DP

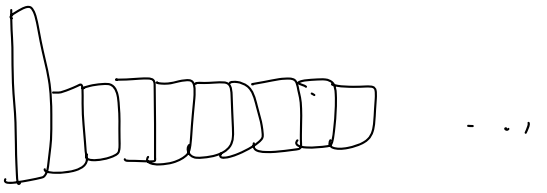
$\Rightarrow 2\varepsilon$-DP

$\Rightarrow (2\varepsilon, \delta)$-DP overall.

Ex: Mode

Spec: want to release element that appears most often in D.

- Obvious approach: histogram, return highest count.
  - Can add $\text{Lap}(1/\varepsilon)$-noise to each count (histogram)
  - Spec: $|X| = k$, 1 element largest, all other elts 2nd largest:



$\Rightarrow$ in expectation, sample element gets noise $\approx \frac{\log k}{\varepsilon}$

$\Rightarrow$ only works if gap b/w largest, 2nd largest $\geq \frac{\log k}{\varepsilon}$ — k can be huge!

Try using PTR instead.

- Repeat $\beta = 0$
- Let γ = diff. b/w largest and 2nd-largest count
- Let $\hat{\gamma} = \gamma + \text{Lap}(1/\varepsilon)$
- If $\hat{\gamma} \leq \frac{1}{\varepsilon}\ln(1/\delta)$, return $\perp$
- else return most frequent elt

Thm: (ε, δ)-DP

Pf: PTR w/ $\beta = 0$!

Thm: Returns the mode w/ prob. $\geq 1-\delta$ as long as diff b/w largest and 2nd largest count $\geq 2\frac{1}{\varepsilon}\ln\frac{1}{\delta}$

Pf: Only fails if $\hat{\gamma} \leq \frac{1}{\varepsilon}\ln\frac{1}{\delta}$

$$\Rightarrow -Lap\left(\tfrac{1}{\varepsilon}\right) \geq \gamma - \tfrac{1}{\varepsilon}\ln\tfrac{1}{\delta} \geq \tfrac{1}{\varepsilon}\ln\tfrac{1}{\delta}$$

$$Pr\left[-Lap\left(\tfrac{1}{\varepsilon}\right) \geq \tfrac{1}{\varepsilon}\ln\tfrac{1}{\delta}\right] \leq e^{-\ln\frac{1}{\delta}} = \delta \quad \checkmark$$

Generalize: "stable histograms".

Laplace histograms: some entry off by $\Theta\left(\frac{\log k}{\varepsilon}\right)$

k could be huge: arbitrarily larger than n: $|\mathcal{D}|$!

Even infinite!

Want to set $\sim\left(\frac{1}{\varepsilon}\ln\frac{1}{\delta}\right)$.

Think of $\delta \leq \frac{1}{n^{\text{w:lol}}}$, so looking for $\frac{1}{\varepsilon}\ln n$ error.

Idea: would be done if only released nonzero counts, but would violate privacy

- only release counts above a threshold, round all others down to 0!

SH:

- For each $x \in X$:
  - if $D_x = 0$, set $a_x = 0$ (D_x = count of x in D)
  - if $D_x > 0$:
    - Set $a_x = D_x + Lap(\frac{1}{\varepsilon})$
    - If $a_x < \frac{1}{\varepsilon} \ln \frac{1}{\delta} + 1$, set $a_y = 0$

Thm: For $\delta \le \frac{1}{n}$, w.h.p. $|a_x - D_x| \le O(\frac{1}{\varepsilon} \ln \frac{1}{\delta})$ $\forall x \in X$

Pf: Definitely true if $D_x = 0$

else, w.h.p. $D_x + Lap(\frac{1}{\varepsilon}) \le D_x + O(\frac{1}{\varepsilon} \log n)$

$\le D_x + O(\frac{1}{\varepsilon} \ln \frac{1}{\delta})$ $\forall x \in X$

$\Rightarrow |a_x - D_x| \le O(\frac{1}{\varepsilon} \ln \frac{1}{\delta}) + \frac{1}{\varepsilon} \ln \frac{1}{\delta} + 1$

$\le O(\frac{1}{\varepsilon} \ln \frac{1}{\delta})$ w.h.p.

Thm: (ε, δ)-DP.

Pf: Let $D \sim D'$ where $D_x = D'_x + 1$ (added x to D)

Only diff. b/w $SH(D)$ and $SH(D')$ is in a_x (all other values identically distributed)

Case 1: $D'_x > 0$

$\Rightarrow$ add $Lap(\frac{1}{\varepsilon})$-noise to both D'_x and D_x

$\Rightarrow$ ε-DP (truncation is post-processing)

Case 2: $D'_x = 0 \Rightarrow a'_x = 0$

$\Rightarrow D_x = 1$

$\Rightarrow Pr[a_x \neq 0] = Pr[Lap(\frac{1}{\varepsilon}) > \frac{1}{\varepsilon} \ln \frac{1}{\delta}] \leq \delta$

$\Rightarrow (0, \delta)$-DP

$\Rightarrow (\varepsilon, \delta)$-DP overall.