1/29/25:

Def: ℓ_1-sensitivity, global sensitivity: $f: \mathcal{D} \to \mathbb{R}^k$

$$\Delta f = \max_{D, D' \text{ neighboring}} \|f(x) - f(y)\|_1$$

How much can a single individual's data change f?

Common case: $k=1$, so $f: \mathcal{D} \to \mathbb{R}$

Ex: Counting queries.

"How many people in this database satisfy property P"?

$\Delta f = 1$

"What fraction": $\Delta f = \frac{1}{n}$

⋮

Can be huge: if data is $\mathbb{R}$, if query is average, max, etc., then Δf unbounded!

Laplace mechanism: add Laplace noise to each entry, scaled by $\Delta f/\varepsilon$. Formally:

Def: Laplace Mechanism: Given $f: \mathcal{D} \to \mathbb{R}^k$, **Laplace mechanism** is $M_L(D, f, \varepsilon) = f(D) + (Y_1, \ldots, Y_k)$, where Y_i's are i.i.d. random vars from $Lap\left(\frac{\Delta f}{\varepsilon}\right)$

Equivalently: $M_L(D, f, \varepsilon) = (Z_1, \ldots, Z_k)$ where Z_i's are i.i.d. random vars from $Lap\left(f(D)_i, \frac{\Delta f}{\varepsilon}\right)$

Ex: counting queries

1 query: $\Delta f = 1 \Rightarrow$ add $Lap(1/\varepsilon)$ noise

k queries: think of f as vector of answers

to all queries

$\Rightarrow \Delta f$ could be k

$\Rightarrow$ add $Lap(k/\varepsilon)$ to **every** query / component!

Ex: histogram queries

Divide range into cells, report count in each cell.

Ex: height. 5'-5'2", 5'2"-5'4", etc.

counting query in each bucket

$\Rightarrow \Delta f = 1$

$\rightarrow$ add $Lap(1/\varepsilon)$ to each query, even though k queries

Q: How good is this mechanism (accuracy/utility/approx).

Fact: If $X \sim Lap(b)$, then

$$Pr[|X| \geq t \cdot b] = \exp(-t) \quad \text{(tail bound)}.$$

Thm: Let $f: \mathcal{D} \to \mathbb{R}^k$, let $y = M_L(D, f, \varepsilon)$.

Then $\forall \delta \in (0,1]$:

$$Pr\left[\|f(D) - y\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \cdot \frac{\Delta f}{\varepsilon}\right] \leq \delta$$

Pf: $\Pr[\|f(D) - y\|_\infty \geq \ln\left(\frac{k}{\delta}\right) \cdot \frac{\Delta f}{\varepsilon}]$

$= \Pr\left[\max_{i \in [k]} |Y_i| \geq \ln\left(\frac{k}{\delta}\right) \cdot \frac{\Delta f}{\varepsilon}\right]$

Y_i = noise added to coordinate i

$\leq k \cdot \Pr\left[|Y_i| \geq \ln\left(\frac{k}{\delta}\right) \cdot \frac{\Delta f}{\varepsilon}\right]$ (union bound, all Y_i's i.i.d.)

$= k e^{-\ln\left(\frac{k}{\delta}\right)}$ (previous fact)

$= \delta$

Ex: First names.

Given list of 10000 names, how many people from last census had each name?

Histogram query, $\Delta f = 1$

$\Rightarrow$ add $\text{Lap}(1)$ noise to each count, get 1-DP

Set $\delta = 0.05$:

with probability 95%, no estimate off by more than $\ln\left(\frac{10000}{0.05} \cdot 1\right) \approx 12.2$.

Pretty good!

Gaussian Mechanism: will talk about more ways get to RDP/CDP/zCDP.

Mechanism: add $N(0, \left(\frac{1}{\varepsilon}\Delta f\right)^2 \ln\frac{1}{\delta})$ noise

↑ variance

Δf: can even use ℓ_2-sensitivity!

$\max_{D,D' \text{ neighboring}}$

Thm: For $\varepsilon \in (0,1)$, this is (ε, δ)-DP.

See Appendix A.

Differentially private Selection:

- In first names example, sps want to know which name is most popular, not actual counts?
- Releases less information: can we be more accurate?
- Generalize: sps k counting queries (not histogram).
- Laplace: add $Lap(\frac{|c|}{\varepsilon})$ noise to each

Report Noisy Max:

- Add $Lap(1/\varepsilon)$ to each, release which is largest.

Thm: RNM is ε-DP

pf: - Let $D = D' \cup \{a\}$.

- Let $c \in \mathbb{N}^{|C|}$ true counts for D, $c' \in \mathbb{N}^{|C|}$ for D'

- Fix $i \in C(|C|)$. WTS: $Pr[i \mid D] \in [e^{-\varepsilon} Pr[i \mid D'], e^{\varepsilon} Pr[i \mid D'])$

- Fix r_{-i}, a draw from $(Lap(1/\varepsilon))^{|C|-1}$ used for all noisy counts other than i.

First: Show $Pr[i \mid D, r_{-i}] \leq e^{\varepsilon} Pr[i \mid D', r_{-i}]$

Let $v^* = \min_{r_i} : c_i + r_i > c_j + r_j \ \forall j \neq i$

(min value of r_i so RNM returns i when other noises fixed)

$\forall j \neq i$:

$$c_i' + (v^* + 1) \geq^{\text{counting}} c_i + v^* \geq^{\text{def}} c_j + r_j$$

$$\geq c_j' + r_j \quad (\text{def of } i, c)$$

So if $r_i \geq v^* + 1$, will return i in D', r_{-i}

$\Rightarrow \Pr[i \mid D', r_{-i}] \geq \Pr[r_i \geq r_i^* + 1]$

$= e^{-\varepsilon} \Pr[r_i \geq r_i^*]$ (Laplace distribution w/ parameter $1/\varepsilon$)

$= e^{-\varepsilon} \Pr[i \mid D, r_{-i}]$

$\Rightarrow \Pr[i \mid D, r_{-i}] \leq e^{\varepsilon} \Pr[i \mid D', r_{-i}]$ ✓

Second: Show $\Pr[i \mid D', r_{-i}] \leq e^{\varepsilon} \Pr[i \mid D', r_{-i}]$

Let $r^* = \min_{r_i} : c'_i + r_i > c'_j + r_j \quad \forall j \neq i$

$\forall j \neq i$:

$c_i + (r^* + 1) \geq c'_i + (r^* + 1)$ (def of c, c')

$> c'_j + (r_j + 1)$ (def of r^*)

$\geq c_j + r_j$ (counting query)

$\Rightarrow$ If $r_i \geq r^* + 1$, will return i in D, r_{-i}

$\Rightarrow \Pr[i \mid D, r_{-i}] \geq \Pr[r_i \geq r^* + 1]$

$= e^{-\varepsilon} \Pr[r_i \geq r^*]$ (Lap($\frac{1}{\varepsilon}$))

$= e^{-\varepsilon} \Pr[i \mid D', r_{-i}]$

$\Rightarrow \Pr[i \mid D', r_{-i}] \leq e^{\varepsilon} \Pr[i \mid D, r_{-i}]$ ✓