Lecture 9

- Selection Problem

  Exponential Mechanism.

Watch out HW2.

Released today or tomorrow.

Previously on Differential Privacy.

Algorithms

→ Randomized Response

→ Laplace Mechanism

↓ generalization

Today

→ Exponential Mechanism

Problem

→ Release Numeric Statistics (avg, histogram, ...)

generalization

→ Selection Problem.

DP Properties

Composition

Post-Processing

Group Privacy.

(Agnostic to ALGs)

Selection Problem

Heavy Hitter

Example. A set of websites $\{1,\dots,d\}$

Each user submits $X_i \subseteq \{1,\dots,d\}$ ← Visit a subset of websites

Winner: website with the highest score : $\forall j \in \{1,\dots,d\}$

$$q(j;x) = \Big|\underbrace{\{ i \mid j \in X_i \}}_{\text{The set of users who visited website } j \text{ in the data set } x.}\Big|$$

Notation

$|S|$ ← Cardinality or how many elements are in S.

↑ Set

Want to find website j such that

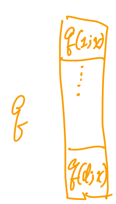
$$\text{Error} = \max_{j^*} q(j^*;x) - q(j;x)$$

is small.

One Proposal:

→ Run Laplace Mechanism to release $q(j;x)$ for all $j \in \{1,\dots,d\}$

→ then output $\hat{j}$ with the maximum noisy score.



$GS_q = d.$

⇒ Adding Laplace noise scaling w/ d.

Example 2: Pricing a digital good.

- Selling an app; what price?
- n people's valuations: $\underbrace{\text{"How much are they willing to pay?"}}_{x_i}$ $:=$ private value.

price

$$\text{Revenue: } \quad g(\overset{\downarrow}{p}; x) = p \cdot \#\{i : x_i \geq p\}.$$

$$\text{Error: } \quad \underbrace{\max_p g(p; x)}_{\text{optimal Revenue.}} - g(A(x), x)$$

$A(x)$ ↑ Price returned by your algorithm A.

4 people:

$x_1 = 1$

$x_2 = 1$

$x_3 = 1$

$x_4 = 4.01$ ← optimal price.

Formulation: Selection Problem

$\mathcal{Y}$: possible outcomes (e.g. websites, prices).

$q : \underline{\mathcal{Y}} \times \underline{\mathcal{X}^n} \longrightarrow \mathbb{R}$ "score" function (e.g., # hits, revenue)

measures how good y is on dataset x.

q is Δ-sensitive if $\forall\, y \in \mathcal{Y}$

$q(y;\cdot)$ has $GS_q \le \Delta$.

Exponential Mechanism. $A_{EM}(x, q, \varepsilon, \Delta)$

Output an outcome y with probability $\boxed{\text{proportional to}}$

$$\exp\left(\frac{\varepsilon}{2\Delta}\, q(y;x)\right).$$

For this class, assume outcome space $\mathcal{Y}$ is finite.

$$\mathbb{P}[A_{EM}(x, q, \varepsilon, \Delta) = y] = \frac{1}{C_x} \cdot \exp\left(\frac{\varepsilon}{2\Delta}\, q(y;x)\right)$$

"Normalization factor" $C_x = \sum_{y' \in \mathcal{Y}} \exp\left(\frac{\varepsilon}{2\Delta}\, q(y';x)\right).$

Privacy Proof.

Theorem. For every Δ-sensitive q, $A_{EM}(\cdot, q, \varepsilon, \Delta)$ is ε-DP.

Proof. Fix any neighbors x & x', any outcome $y \in Y$.

Goal: to show

$\mathbb{P}[A(x)=y]$ ← Plug in

$\leq e^{\varepsilon} \mathbb{P}[A(x')=y]$

↑ Plug in

$$\frac{\mathbb{P}[A(x)=y]}{\mathbb{P}[A(x')=y]} = \frac{\frac{1}{C_x} \cdot \exp\left(\frac{\varepsilon}{2\Delta} q(y;x)\right)}{\frac{1}{C_{x'}} \cdot \exp\left(\frac{\varepsilon}{2\Delta} q(y;x')\right)}$$

$$= \underbrace{\frac{C_{x'}}{C_x}}_{\leq \exp\left(\frac{\varepsilon}{2}\right)?} \cdot \underbrace{\frac{\exp\left(\frac{\varepsilon}{2\Delta} q(y;x)\right)}{\exp\left(\frac{\varepsilon}{2\Delta} q(y;x')\right)}}_{\exp\left(\frac{\varepsilon}{2\Delta}\left(\underbrace{q(y;x)-q(y;x')}_{\leq \Delta}\right)\right) \leq \exp\left(\frac{\varepsilon}{2}\right)} \leq \exp(\varepsilon)$$

$$C_{x'} = \sum_{y' \in Y} \exp\left(\frac{\varepsilon}{2\Delta} q(y';x')\right)$$

$$\leq \sum_{y' \in Y} \exp\left(\frac{\varepsilon}{2}\right) \cdot \exp\left(\frac{\varepsilon}{2\Delta} q(y';x)\right)$$

$$= \exp\left(\frac{\varepsilon}{2}\right) \cdot C_x$$

$$\Longrightarrow \frac{C_{x'}}{C_x} \leq \exp\left(\frac{\varepsilon}{2}\right).$$

For any y'

$$\exp\left(\frac{\varepsilon}{2\Delta} q(y';x)\right)$$

$$\leq \exp\left(\frac{\varepsilon}{2}\right) \cdot \exp\left(\frac{\varepsilon}{2\Delta} q(y';x')\right)$$

because

$$\frac{\exp\left(\frac{\varepsilon}{2\Delta} q(y';x)\right)}{\exp\left(\frac{\varepsilon}{2\Delta} q(y';x')\right)} \leq \exp\left(\frac{\varepsilon}{2}\right)$$

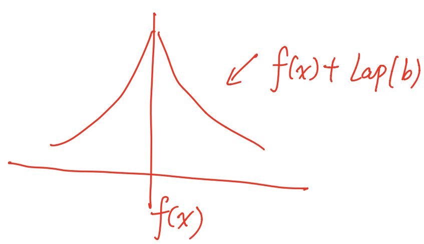
Exp. Mechanism is everywhere.

Laplace Mechanism. $Y \in \mathbb{R}$

$f : \mathcal{X}^n \mapsto \mathbb{R}$, $\quad q(y; x) = -\underbrace{|y - f(x)|}_{\text{error}}$

Laplace mechanism sample y w.p.

proportional to $\quad \exp\left(\frac{-\varepsilon}{2 GS_f} |y - f(x)|\right)$



Randomized Response $\quad Y = \{0,1\}^n$

$$q(y; x) = \|y - x\|_1$$

private bits

RR samples y with prob proportional to

$$\exp\left(\frac{-\varepsilon}{2} \|y - x\|_1\right).$$

Recitation on Friday (in-person) w/ Justin.

How useful is EM?

Theorem. ($\mathcal{Y}$ is finite) Let $\mathcal{Y} = [d]$

Then.
$$\mathop{\mathbb{E}}_{Y \sim A_{EM}} \left[q_{max}(x) - q(Y; x) \right] \leq \frac{2\Delta}{\varepsilon} \left(\ln(d) + 1 \right)$$

"Tail Bound"
$\forall t > 0,$
$$\mathop{\mathbb{P}}_{Y \sim A_{EM}} \left[q_{max}(x) - q(Y; x) \geq \frac{2\Delta}{\varepsilon} \left(\ln(d) + t \right) \right] < e^{-t}$$

Proof.