Lecture 10

- Selection Problem Continued

  Exponential Mechanism.

Announcement:

→ HW2 due in two weeks 10/17 (Sunday)

→ HW1 Solution will be posted.

→ Next week No Class on 10/13 Weds.
Optional Bonus Guest lecture on Zoom. (?)
Make-up class on 10/15 Friday.
during Recitation Time.

Selection Problem

Heavy Hitter

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

Each user submits $X_i \subseteq \{1, \dots, d\}$

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

$$q(j; x) = |\{i \mid j \in X_i\}|.$$

Proposed solution: ① Laplace Mech. + Report the noisy max.

GS := d.

② Exponential Mech.

Example 2: Pricing a digital good.

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

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

Error: $\underbrace{\max_p f(p; x)}_{\text{Optimal Revenue.}} - f(\underset{\substack{\uparrow \\ \text{Price returned by} \\ \text{your algorithm } A.}}{A(x)}, x)$

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 proportional to

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

For this class, assume outcome space 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" $C_x = \sum_{y' \in Y} \exp\left(\frac{\varepsilon}{2\Delta} q(y'; x)\right).$

Privacy Proof.

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

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

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

$$\frac{\mathbb{P}[A(x)=y]}{\mathbb{P}[A(x')=y]} = \underbrace{\frac{C_{x'}}{C_x}}_{\leq e^{\varepsilon/2}} \cdot \underbrace{\frac{\exp\left(\frac{\varepsilon}{2\Delta} q(y;x)\right)}{\exp\left(\frac{\varepsilon}{2\Delta} q(y;x')\right)}}_{\leq e^{\varepsilon/2}}$$

show

Claim: For any $y' \in Y$,

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

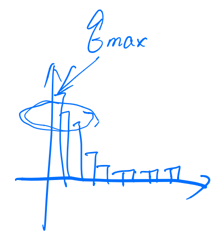
$\Rightarrow$ Set $y' = y$: $$\frac{\exp\left(\frac{\varepsilon}{2\Delta} q(y;x)\right)}{\exp\left(\frac{\varepsilon}{2\Delta} q(y;x')\right)} \leq e^{\varepsilon/2}$$

$$\Rightarrow \quad C_{x'} = \sum_{y'} \exp\left(\frac{\varepsilon}{2\Delta} q(y';x')\right)$$
$$\leq \sum_{y'} \exp\left(\frac{\varepsilon}{2\Delta} q(y';x)\right) \cdot e^{\varepsilon/2}$$
$$\leq C_x \cdot e^{\varepsilon/2}$$

How useful is EM?

Compare $q_{max}(x) = \max_y q(y;x)$

$$\text{error} := \underbrace{q_{max}(x) - q(A_{EM}(x);x)}_{\text{minimize}}$$



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

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

$$\forall t > 0, \quad \mathbb{P}_{\hat{y} \sim A_{EM}} \Big[\underbrace{q_{max}(x) - q(\hat{y};x)}_{\text{error}} \ge \underbrace{\tfrac{2\Delta}{\varepsilon} \left(\ln(d) + t \right)}_{\text{threshold / error Bound}} \Big] < \underbrace{e^{-t}}_{\text{"failure probability"}} \qquad 1\%$$

Tail Bound on the error.

Proof. Fix any data set $x \in \mathcal{X}^n$.

$$B_t = \Big\{ y \in \mathcal{Y} \;\Big|\; q(y;x) < \underbrace{q_{max} - \tfrac{2\Delta}{\varepsilon}(\ln(d)+t)}_{\text{err bound}} \Big\}$$

"Bad set of outcomes"

plug in

For any $y \in B_t$,

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

$$< \frac{1}{C_x} \exp\left(\frac{\varepsilon}{2\Delta} \left(q_{max} - \frac{2\Delta}{\varepsilon} (\ln(d) + t) \right) \right)$$

$$= \underbrace{\frac{1}{C_x} \exp\left(\frac{\varepsilon}{2\Delta} q_{max} \right)}_{\mathbb{P}[A(x) = Best] \le 1} \underbrace{\exp\left(-\ln(d) \right)}_{\frac{1}{d}} \cdot \exp(-t)$$

Our goal: $\mathbb{P}[A(x) \in B_t] < e^{-t}$

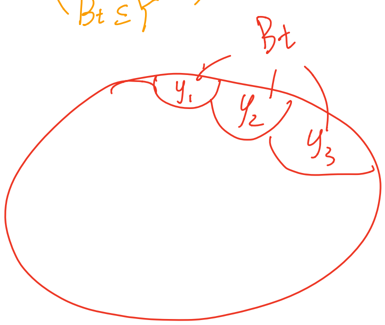
$$\mathbb{P}[A(x) \in B_t] \le \frac{\mathbb{P}[A(x) \in B_t]}{\mathbb{P}[A(x) = Best]}$$

$(\mathbb{P}[A(x) = Best] \le 1)$

$$(\text{Union Bound}) < \frac{|B_t| \cdot \mathbb{P}[A(x) = Best] \cdot \frac{1}{d} \cdot \exp(-t)}{\mathbb{P}[A(x) = Best]}$$

Upper Bound on prob. for each $y \in B_t$

$$\begin{pmatrix} |B_t| \le d \\ B_t \subseteq Y \end{pmatrix} \le d \cdot \frac{1}{d} \exp(-t) = \exp(-t).$$



Selection Problem Example

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

Each user submits $x_i \subseteq \{1, \dots, d\}$

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

$$q(j; x) = \left| \{ i \mid j \in x_i \} \right|.$$ "# users visiting"

$$\text{Error} = \underbrace{\max_j q(j;x)}_{q_{max}} - q(A(x); x).$$

Error Bounds:

Set $t = \ln\left(\frac{1}{100}\right)$

① Exp Mech.

$$\mathbb{P}_{\hat{y} \sim A_{EM}}\left[\underbrace{q_{max}(x) - q(\hat{y}; x)}_{\text{error}} \geq \underbrace{\frac{2\Delta}{\varepsilon}\left(\ln(d) + t\right)}_{\text{threshold / error Bound}}\right] < \underbrace{e^{-t}}_{\text{"failure probability"}} \quad e^{-t} = 1\%$$

W. p. 99%, error < $\boxed{\dfrac{2 \ln(100d)}{\varepsilon}}$

$\forall y \in Y$, $\left| q(y, x) - q(y, x') \right| \leq \Delta$, $\Delta = 1$.

② Laplace Mech.

Release $q(y; x) + \text{Lap}\left(\frac{GS_q}{\varepsilon}\right) \quad \forall y.$

$$GS_q = \max_{x, x'} \sum_{j=1}^{d} \left| q(j; x) - q(j; x') \right| = d.$$

$$\text{Error} \geq \left(\frac{d}{\varepsilon}\right)$$