Lecture 7

Recap and continue

- Solving Seletion Problem with DP.
  - Exponential Mechanism <- (General)
  - Report Noisy Max
- (If have time): Continual Release

Logistics: HW Posted on Canvas

Discussion On Canvas; Due 2/28 Sunday

On schedule page of the course webpage → Lecture notes & Slides are posted.

Selection Problem Example

Heavy Hitter.

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

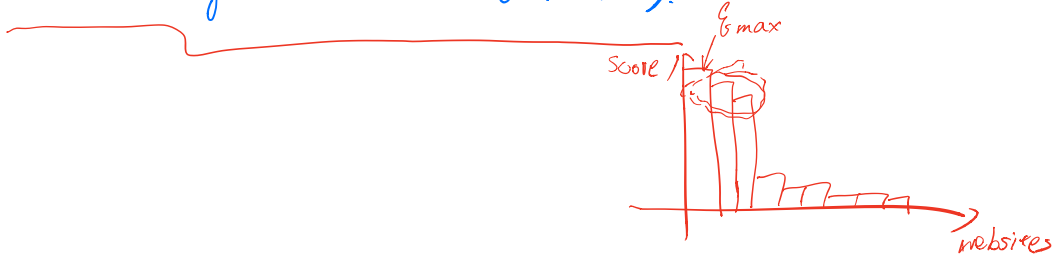
Each user submits $x_i \subseteq \{1, \dots, d\}$

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

$$q(j; x) = \left| \{ i \mid j \in x_i \} \right|$$

"# users visiting"

$$\text{Error} = \max_j q(j; x) - q(A(x); x)$$



Formulation: Selection Problem

Y: possible outcomes

$q : Y \times X^n \to \mathbb{R}$ "score" function

(Y: outcome, X^n: dataset) measures how good y is on dataset x.

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

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

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

Output an outcome y with prob. $\propto \exp\left(\frac{\varepsilon}{2\Delta} q(y;x)\right)$

↑ proportional to

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

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

↑ normalization

C_x needs to be well-defined for all x when Y is infinite / continuous.

Privacy Proof.

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

Proof. First focus on Y being finite.

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

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

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

$$= \exp\left(\frac{\varepsilon}{2\Delta}\left(q(y;x) - q(y;x')\right)\right) \qquad \frac{C_{x'}}{C_x} \leq e^{\frac{\varepsilon}{2}}$$

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

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

$$= e^{\frac{\varepsilon}{2}} \cdot C_x$$

(Extend this to any $E \subseteq Y$)

Continuous / Infinite Y is the same: but require $\forall x$, C_x to be well-defined.

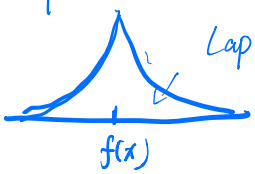
Exp. Mechanism is everywhere.

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

$f: \mathcal{X}^n \mapsto \mathbb{R}$, $g(y; x) = -|y - f(x)|$ $\left(-\|y - f(x)\|_1\right)$

Exp. mech. sample y w.p. $\propto \exp\left(\frac{-\varepsilon}{2GS_f} \cdot \left(-|y - f(x)|\right)\right)$

Lap. mech. sample y w.p. $\propto \exp\left(\frac{-\varepsilon}{GS_f}\left(-|y - f(x)|\right)\right)$



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

$g(y; x) = -\|y - x\|_1$ "measure agreement"

Exp. mech. sample y w.p $\propto \exp\left(\frac{-\varepsilon}{2}\left(-\|y - x\|_1\right)\right)$

RR ——— $\propto \exp\left(\frac{-\varepsilon}{1}\left(-\|y - x\|_1\right)\right)$

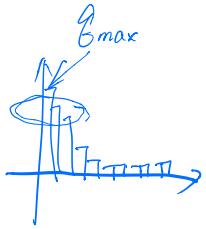
$\forall$ any ε-DP A,

You can view it as an exponential Mech. (w/ a factor of 2 in ε)

How useful is EM?

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

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



Theorem. (Y is finite) Let $Y=[d]$

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

"Tail Bound"

→ $\forall t>0$, $\mathop{\mathbb{P}}_{Y\sim A_{EM}}\left[\underbrace{q_{max}(x) - q(Y;x) \ge \boxed{\frac{2\Delta}{\varepsilon}\left(\ln(d)+t\right)}}_{\text{"Large error" event.}}\right] < \underbrace{\boxed{e^{-t}}}_{\text{"failure probability"}}$

err bdd. ↙ 1% ↙

Proof. Fix $x \in X^n$. $B_t = \{y : q(y) < q_{max} - \frac{2\Delta}{\varepsilon}(\ln(d)+t)\}$

For any $y \in B_t$: $\mathbb{P}[Y=y] < \frac{1}{C_x}\exp\left(\frac{\varepsilon}{2\Delta}\left(q_{max} - \frac{2\Delta}{\varepsilon}(\ln(d)+t)\right)\right)$

(single y)

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

→ $\mathbb{P}[Y \in B_t] \le \frac{\mathbb{P}[Y\in B_t]}{\mathbb{P}[q(Y)=q_{max}]} \le \frac{|B_t| \cdot \frac{1}{C_x}\exp\left(\frac{\varepsilon}{2\Delta}q_{max}\right)\cdot\frac{1}{d}\exp(-t)}{\frac{1}{C_x}\exp\left(\frac{\varepsilon}{2\Delta}q_{max}\right)}$

$(|B_t| \le d-1) \quad < d\cdot\frac{1}{d}\cdot\exp(-t) = e^{-t}$.

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: $\forall j \in \{1, \dots, d\}$

$$f(j; x) = \left| \{ i \mid j \in x_i \} \right|.$$

"# users visiting"

$$\text{Error} = \max_j f(j; x) - f(A(x); x)$$

Exp. mech. finds a website w/ score

$$\geq f_{max} - \frac{2 \ln(100d)}{\varepsilon}$$

w.p. 99%

If release all scores w/ Laplace mech.

$$\text{err} \approx \frac{d}{\varepsilon}.$$

Report Noisy Max

Y: possible outcomes

$q: Y \times X^n \to \mathbb{R}$ "score" function

measures how good y is on dataset x.

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

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

$A_{RNM}(x, q, \Delta, \varepsilon)$:

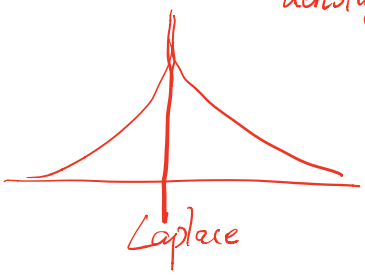
1) For $y = 1, \dots, d$:

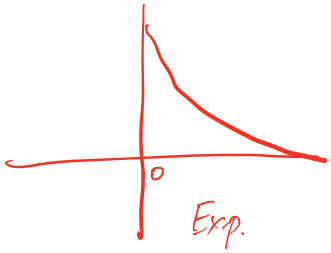
$\tilde{q}(y) = q(y;x) + Z_y$, $Z_y \sim Exp\left(\frac{2\Delta}{\varepsilon}\right)$ ← "Exponential Distribution"

2) Return $\hat{y} = \arg\max_y \tilde{q}(y)$.

$Exp(\lambda)$: distribution over $[0, \infty)$

density $h_\lambda(z) = \frac{1}{\lambda} \exp\left(\frac{-z}{\lambda}\right)$





Theorem. A_{RNM} is ε-DP.

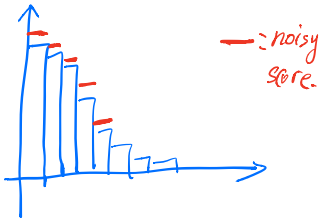
$\forall x$, $y = A_{RNM}(x, q, \Delta, \varepsilon)$

Expectation: $\mathbb{E}[q(y;x)] \geq q_{max} - \frac{2\Delta}{\varepsilon}[\ln(d)+1]$

Tail: $\forall t>0$, $\mathbb{P}[q(y;x) < q_{max} - \frac{2\Delta}{\varepsilon}(\ln(d)+t)] < \exp(-t)$

Utility.

Suffices to bound the max noise.



If $z_1, \dots, z_d \sim Exp(\lambda)$

$z^* = \max_{j \in [d]} z_j$.

a) $\mathbb{E}[z^*] < \lambda(\ln(d)+1)$

b) $\mathbb{P}[z^* \geq \lambda(\ln(d)+t)] < e^{-t}$

Proof. 2b) $z^* \geq \lambda(\ln(d)+t) \iff \exists j,\ z_j \geq \lambda(\ln(d)+t)$.

$$\mathbb{P}[\exists j,\ z_j \geq \lambda(\ln(d)+t)] \underset{\text{Union Bound}}{\leq} \sum_{j\in[d]} \mathbb{P}[z_j \geq \lambda(\ln(d)+t)]$$

$$= d \cdot e^{-\ln(d)-t}$$

$$= e^{-t}.$$

← Prob. Space

2a) $\mathbb{E}[z^*] = \int_0^\infty \mathbb{P}[z^* > z]\, dz$

$$= \int_0^{\ln(d)} \underbrace{\mathbb{P}[z^* > z]}_{\leq 1}\, dz + \int_{\ln(d)}^\infty \underbrace{\mathbb{P}[z^* > z]}_{\leq e^{-t}\text{ tail bound.}}\, dz$$

$$= \ln(d) + \int_0^\infty e^{-t} = \ln(d) + 1.$$

Other Perturbations

→ Gumbel noise. $h(z) = e^{-(z + e^{-z})}$

$\Leftrightarrow$ Exp. mechanism.

"Gumbel max trick"

→ Laplace ~~Dist~~ Noise.

Example 2: Pricing a digital good.

- Selling an app; what price?
- n people's valuations: "How much are they willing to pay?" $= x_i$

Revenue:

$$f(p; x) = p \cdot \#\{i : x_i \geq p\}$$

Error

$$\max_p f(p, x) - f(A(x), x)$$