Differential Privacy: Exponential Mechanism

CompSci 590.03
Instructor: Ashwin Machanavajjhala
Limitations of output perturbation

• What if the answer is non-numeric?
  – “what is the most common nationality in this room”: Chinese/Indian/American...
  – Other examples?

• What if the perturbed answer is not as good as the real answer?
  – “Which price would bring the most money from a set of buyers?”
Example: Items for sale

- If price is set at $100, make a revenue of $400
- If price is set at $401, make a revenue of $401
- Best price: $401, Next best: $100
- Revenue at $402 = $0
- Revenue at $101 = $101
Exponential Mechanism

• Consider some algorithm A (can be deterministic or probabilistic):

• How to construct a differentially private version of A?
Exponential Mechanism

• Construct a scoring function $w: \text{Inputs} \times \text{Outputs} \rightarrow R$

Examples:
• $w(D, O) = c$, for all $D \in \text{Inputs}$ and $O \in \text{Outputs}$.
• $w(D,O) = P[A(D) = O]$, for all $D \in \text{Inputs}$ and $O \in \text{Outputs}$.

• For good utility $w(D,O)$ should mirror the true algorithm as well as possible.
Exponential Mechanism

• Construct a scoring function \( w: \text{Inputs} \times \text{Outputs} \rightarrow R \)

• Sensitivity of \( w \)

\[
\Delta_w = \max_{D, D', O} |w(D, O) - w(D', O')|
\]

where \( D, D' \) differ in one tuple
**Exponential Mechanism**

*Algorithm $E_w^\varepsilon(D)$*

- Given an input $D$, and a scoring function $w$,
  Randomly sample an output $O$ from $Outputs$ with probability

$$
\frac{\varepsilon \cdot w(D, O)}{\sum_{Q \in Outputs} e^{\frac{\varepsilon}{2\Delta} \cdot w(D, Q)}}
$$

- Note that for every output $O$, probability $O$ is output $> 0$. 

Lecture 10: 590.03 Fall 13
Theorem

Algorithm $E_w^ε(D)$ satisfies $ε$ differential privacy.
Utility of the Exponential Mechanism

• Depends on the choice of scoring function – weight given to the best output.

• E.g.,
  “What is the most common nationality?”
  \[ w(D,\text{nationality}) = \# \text{ people in D having that nationality} \]

  Sensitivity of \( w \) is 1.

• Q: What will the output look like?
Utility of Exponential Mechanism

- Let $OPT(D) = $ nationality with the max score
- Let $O_{OPT} = \{O \in \text{Outputs} : w(D,O) = OPT(D)\}$

- Let the exponential mechanism return an output $O^*$

Theorem:

$$\Pr \left[ w(D, O^*) \leq OPT(D) - \frac{2\Delta}{\varepsilon} \left( \log \frac{|\text{Outputs}|}{|O_{OPT}|} + t \right) \right] \leq e^{-t}$$
Utility of Exponential Mechanism

Theorem:

$$\Pr \left[ w(D, o^*) \leq OPT(D) - \frac{2\Delta}{\varepsilon} \left( \log \frac{|Outputs|}{|O_{OPT}|} + t \right) \right] \leq e^{-t}$$

Suppose there are 4 nationalities

Outputs = {Chinese, Indian, American, Greek}

Exponential mechanism will output some nationality that is shared by at least \(K\) people with probability \(1-e^{-3}(=0.95)\), where

$$K \geq OPT - 2(\log(4) + 3)/\varepsilon = OPT - 6.8/\varepsilon$$
Laplace versus Exponential Mechanism

• Let \( f \) be a function on tables that returns a real number.

• Define: score function \( w(D, O) = |f(D) - O| \)

• Sensitivity of \( w = \max_{D, D'} (|f(D) - O| - |f(D') - O|) \)
  \[ \leq \max_{D, D'} |f(D) - f(D')| = \text{sensitivity of } f \]

• Exponential mechanisms returns an output \( f(D) + \eta \) with probability proportional to

\[ e^{\frac{\varepsilon}{2|D - O|}} |f(D) - f(D) - \eta| \quad \text{Laplace noise with parameter } 2\Delta/\varepsilon \]
Summary of Exponential Mechanism

• Differential privacy for cases when output perturbation does not make sense.

• Idea: Make better outputs exponentially more likely; Sample from the resulting distribution.

• Every differentially private algorithm is captured by exponential mechanism.
  – By choosing the appropriate score function.
Summary of Exponential Mechanism

• Utility of the mechanism only depends on \( \log(|\text{Outputs}|) \)
  – Can work well even if output space is exponential in the input

• However, sampling an output may not be computationally efficient if output space is large.