CS3325: Private Data Analysis **Fall 2023**

Lecture 4: Differentially Private Selection

Lecturer: Di Wang *Scribes: Di Wang*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

4.1 Exponential Mechanism

The Laplace mechanism works well when the computation we want to carry out returns a vector to which we can add noise, and computations global sensitivity is not too high. What happens when when adding noise to the result makes no sense? The exponential mechanism is the natural starting point for designing differentially private algorithms.

Well motivate the mechanism with two problems, both of which have a selection flavor:

Example 1: (Approval voting) Suppose there are d candidates and n voters (users), each voter can vote for as many candidates as they want; the candidate with the most votes win. One can think of each voters input as a subset $x_i \subseteq \{1, 2, \cdots, n\}$. The score of candidate j is the number of voters who included j in their subsets, that is $q(j; D) = |i : j \in x_i|$. The highest-scoring candidates wins.

We wish to run the election differentially privately. We wont necessarily be able to get the exact winner, but maybe we can get a name with almost the maximum number of votes. One approach is to use the Laplace mechanism to release noisy versions of all the scores. But the global sensitivity of the whole list is d and then we would add noise $\frac{d^2}{\epsilon}$ to each to preserve ϵ-DP. Can we obtain the name someone whose score is much closer than $\frac{d^2}{\epsilon}$ to the highest?

Example 2: (Prices of a digital good) Suppose you recorded a song and you want to sell it online. You talk with n people and find out the price $x_i \in [0, 1]$, each person would willing to pay for a download of the song. Assuming that respondents answered truthfully, a reasonable estimate for the revenue you would get from selling the download at price p is $q(p, D) = p \times |i; x_i \geq p|$.

Adding noise to the best price might not make sense: For example, if everyone had the same maximum price $x_i = 0.7$ for you song, the best price for you to charge would be 0.7, Charging 0.69 would also be ok (you would still make nearly as much as possible), but charging 0.71 would result in no one buying your song.

These examples share a common structure. They are both special cases of a general selection problem, specified by:

- A set $\mathcal{Y}$ of possible outputs;
- A score function $q : \mathcal{Y} \times \mathcal{X}^n \mapsto \mathbb{R}$ which measures the "goodness" of each output for a dataset. Given $D \in \mathcal{X}^n$, our goal is to find $y \in \mathcal{Y}$ which approximately maximize $q(y; D)$.
- A sensitivity bound $\Delta > 0$ such that $q(y; \cdot)$ is Δ-sensitive for every y, that is

$$\sup_{y \in \mathcal{Y}} \sup_{D \sim D'} |q(y; D) - q(y; D')| \leq \Delta. \tag{4.1}$$

Algorithm 1 Exponential Mechanism

Require: Assume that the score function $q(y;\cdot) : \mathcal{X}^n \mapsto \mathbb{R}$ is Δ-sensitive for all $y \in \mathcal{Y}$.

1: Select Y from the distribution with $\mathbb{P}(Y = y) \propto \exp(\frac{\epsilon}{2\Delta} q(y; D))$.

2: **return** Y

Given these elements, we get Algorithm 1. The idea is that given the score function $q(\cdot; D)$ that assigns a number to each element $y \in \mathcal{Y}$, we define a probability distribution which generates each element in $\mathcal{Y}$ with probability proportional to $\exp(\frac{\epsilon}{2\Delta} q(y; D))$ that is, we sample elements with a probability 2Δ that grows exponentially with their score.

When is this algorithm even well defined? When $\mathcal{Y}$ is finite the algorithm is well-defined since we can set

$$\mathbb{P}(Y = y) = \frac{\exp(\frac{\epsilon}{2\Delta} q(y; D))}{\sum_{y \in \mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D))}.$$

In fact, the algorithm may be reasonable even if over the infinite domains, and even continuous ones. If $\int_{\mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D))dy$ is well-defined.

Theorem 4.1 *If $q(y;\cdot)$ is Δ-sensitive for every y, then the exponential mechanism is ϵ-differentially private.*

Proof: Assume for simplicity that $\mathcal{Y}$ is finite (otherwise we can use the density function). For any output y and dataset D we have

$$\mathbb{P}(Y = y|D) = \frac{\exp(\frac{\epsilon}{2\Delta} q(y; D))}{\sum_{y \in \mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D))}.$$

For each term we have

$$\frac{\exp(\frac{\epsilon}{2\Delta} q(y; D))}{\exp(\frac{\epsilon}{2\Delta} q(y; D'))} = \exp(\frac{\epsilon}{2\Delta}(q(y; D) - q(y; D'))) \leq \exp(\frac{\epsilon}{2}), \tag{4.2}$$

and similarly for the normalizing constants,

$$\frac{\sum_{y \in \mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D)}{\sum_{y \in \mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D')} \leq \frac{\sum_{y \in \mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D)}{e^{-\frac{\epsilon}{2}} \sum_{y \in \mathcal{Y}} \exp(\frac{\epsilon}{2\Delta} q(y; D)} = \exp(\frac{\epsilon}{2}). \tag{4.3}$$

Thus, in total we have

$$\frac{\mathbb{P}(Y = y|D)}{\mathbb{P}(Y = y|D')} \leq e^{\epsilon}. \tag{4.4}$$

■

We now have a very general tool in our toolbox, which can be used to design an algorithm for any problem where we can assign possible outputs a score according to their desirability. The algorithm is always differentially private. The question is, when is this approach actually useful? Does it help us address approval voting and price selection, the two examples problems we started out with?

Just how useful the exponential mechanism is depends a lot on the exact problem structure. But we can write down a few clean and generally useful bounds. In the non-private case, the best we can hope for from a selection algorithm is that, it outputs a candidate $y^*(D) \in \mathcal{Y}$ with the highest possible score, denoted as

$$q_{\max}(D) = \max_{y \in \mathcal{Y}} q(y; D), y^*(D) = \arg\max_{y \in \mathcal{Y}} q(y; D). \tag{4.5}$$

Theorem 4.2 *Suppose $\mathcal{Y}$ is finite and has size d. Then for every Δ-sensitive score function q, for every dataset D and every $t > 0$, the output of the exponential mechanism satisfies*

$$\mathbb{P}(q(y;D) < q_{\max}(D) - \frac{2\Delta(\ln d + t)}{\epsilon}) \leq e^{-t}. \tag{4.6}$$

In particular we have

$$\mathbb{E}(q(y;D)) \geq q_{\max}(D) - \frac{2\Delta(\ln d + 1)}{\epsilon}. \tag{4.7}$$

Proof: Consider a fixed dataset D and a score function q. For convenience we will drop the D symbol in the score function. We divide the possible outputs into sets G_t and B_t for "good" and "bad" outputs, where

$$G_t = \{y \in \mathcal{Y} : q(y) > q_{\max} - \frac{2\Delta}{\epsilon}(\ln d + t)\}$$
$$B_t = \{y \in \mathcal{Y} : q(y) \leq q_{\max} - \frac{2\Delta}{\epsilon}(\ln d + t)\}$$

To prove the first part, we need to show that $\mathbb{P}(B_t) \leq e^{-t}$. We have

$$\mathbb{P}(B_t) < \frac{\mathbb{P}(B_t)}{\mathbb{P}(Y = y^*)} = \frac{\sum_{y \in B_t} \mathbb{P}(Y = y)}{\mathbb{P}(Y = y^*)} = \frac{\sum_{y \in B_t} \exp(\frac{\epsilon}{2\Delta} q(y))}{\exp(\frac{\epsilon}{2\Delta} q_{\max})}.$$

Since $y \in B_t$ satisfies $q(y) \leq q_{\max} - \frac{2\Delta}{\epsilon}(\ln d + t)$, the sum in the numerator is at most $|B_t| \exp(\frac{\epsilon}{2\Delta}(q_{\max} - \frac{2\Delta}{\epsilon}(\ln d + t)))$ and we get that

$$\mathbb{P}(B_t) < \frac{|B_t| \exp(\frac{\epsilon}{2\Delta}(q_{\max} - \frac{2\Delta}{\epsilon}(\ln d + t)))}{\exp(\frac{\epsilon}{2\Delta} q_{\max})} = |B_t| \exp(-\ln d - t) = \frac{|B_t|}{d} \exp(-t). \tag{4.8}$$

The last part from the fact that $\mathbb{E}(Z) = \int_{z=0}^{\infty} \mathbb{P}(Z > z) dz$. If we denote $Z = \frac{\epsilon}{2\Delta}(q_{\max} - q(Y))$, we have

$$\mathbb{E}(Z) = \int_{z=0}^{\infty} \mathbb{P}(Z > z) dz = \int_{t=-\ln(d)}^{\infty} \mathbb{P}(Z > \ln(d) + t) dt \leq \int_{t=-\ln(d)}^{0} 1 dt + \int_{t=0}^{\infty} \exp(-t) dt = \ln(d) + 1. \tag{4.9}$$

■

Approval Voting: Lets apply our new Proposition to the approval voting example. The scores there are counts, and have sensitivity 1. Let max be the score of the most popular candidate, and suppose $d = 100$a reasonable number of candidate names for the mascotand $\epsilon = 0.5$. The previous theorem shows that with probability at least 0.99, well get a candidate whose score is at most $q_{\max} - \frac{2}{0.5}(\log 100 + \log 1/0.01) \approx q_{\max} - 36.8$. This is much better than the Laplacian mechanism.

4.2 Noisy-Max Mechanism

Algorithm 2 Noisy-Max Mechanism

Require: Assume that the score function $q(y;\cdot) : \mathcal{X}^n \mapsto \mathbb{R}$ is Δ-sensitive for all $y \in \mathcal{Y} = \{1, 2, \cdots, d\}$.

1: Select $Z_1, Z_2, \cdots Z_d \sim \text{Exp}(\frac{2\Delta}{\epsilon})$ i.i.d.;

2: **return** $Y = \arg\max_{y \in \{1,2,\cdots\}}(q(y;D) + Z_y)$;

When the domain is finite, it is often more convenient to work with another algorithm which behaves very similarity to the exponential mechanism. The idea is to add noise with expected magnitude $\frac{2\Delta}{\epsilon}$ to each item's score. independent of the number of possible outputs. The algorithm returns the output with the highest noisy score. The distribution being used to generate noise is the exponential distribution $\text{Exp}(\lambda)$, a distribution over the non-negative real numbers $[0, +\infty)$ with density $h_\lambda(y) = \frac{1}{\lambda} \exp(-\frac{y}{\lambda})$.

Theorem 4.3 *The Noisy-Max mechanism is ϵ-differentially private.*

Proof: Consider the case where $Y = i \in [d]$, we will fist fix the noises $Z_1, \cdots, Z_{i-1}, Z_{i+1}, \cdots, Z_d$ and denote it as Z_{-i}. We will show that for any neighboring datasets $D \sim D'$,

$$\mathbb{P}(Y = i|D, Z_{-i}) \leq e^{\epsilon}\mathbb{P}(Y = i|D', Z_{-i}). \tag{4.10}$$

If so then we can proof the theorem by taking the integral w.r.t Z_{-i}.

To show the above inequality we define

$$z^* = \arg\min_z q(i; D) + z > q(j; D) + Z_j, \forall j \neq i. \tag{4.11}$$

Thus we have given D and fixed Z_{-i}, the output $Y = i$ if and only if $z_i \geq z^*$. Thus,

$$\begin{aligned}
&q(i; D) + z^* \geq q(j; D) + z_j \implies \\
&q(i; D') + \Delta + z^* \geq q(j; D') + z_j - \Delta \\
&q(i; D') + (z^* + 2\Delta) \geq q(j; D') + z_j.
\end{aligned}$$

Thus, for dataset D' with fixed Z_{-i}, if $z_i \geq z^* + 2\Delta$, the output for D' will also be i. Thus we have

$$\begin{aligned}
&\mathbb{P}(Y = i|D', Z_{-i}) \geq \mathbb{P}(z_i \geq z^* + 2\Delta) \\
&= \int_{z^*+2\Delta}^{\infty} \frac{\epsilon}{2\Delta} \exp(-\frac{\epsilon y}{2\Delta})dy = e^{-\epsilon} \int_{z^*}^{\infty} \frac{\epsilon}{2\Delta} \exp(-\frac{\epsilon y}{2\Delta})dy \\
&= e^{-\epsilon}\mathbb{P}(Y = i|D, Z_{-i}).
\end{aligned}$$

■

Next, we will provide the utility of Algorithm 2. First we provide the tail bound of the exponential distributions.

Lemma 4.4 (Tail Bounds for Exponential Distributions) *We have the following properties of exponential distributions:*

- *If $Z \sim Exp(\lambda)$, then $\mathbb{P}(Z \geq t\lambda) \leq e^{-t}$ for all $t \geq 0$.*
- *If $Z_1, Z_2, \cdots, Z_d \sim Exp(\lambda)$ i.i.d., and $Z_{\max} = \max_{i=1}^d Z_i$, then $\mathbb{P}(Z_{\max} > \lambda(\ln(d) + t)) \leq e^{-t}$ for all $t \geq 0$, and $\mathbb{E}(Z_{\max}) \leq \lambda(\ln(d) + 1)$*

Proof: The proof is almost the same as the proof of the properties of the Laplace distributions. ■

Theorem 4.5 *Suppose $\mathcal{Y}$ is finite and has size d. Then for every Δ-sensitive score function q, for every dataset D and every $t > 0$, the output of the Noisy-Max mechanism satisfies*

$$\mathbb{P}(q(y; D) < q_{\max}(D) - \frac{2\Delta(\ln d + t)}{\epsilon}) \leq e^{-t}. \tag{4.12}$$

In particular we have

$$\mathbb{E}(q(y; D)) \geq q_{\max}(D) - \frac{2\Delta(\ln d + 1)}{\epsilon}. \tag{4.13}$$

Proof: By the definition of the output y and since the exponential distribution is non-negative, we have

$$q(y; D) + Z_y > q(y^*; D) + Z_{y^*} \geq q(y^*; D)$$

Thus, with probability at least $1 - e^{-t}$

$$q(y^*; D) - q(y; D) \leq Z_y \leq Z_{\max} \leq \frac{2\Delta}{\epsilon}(\ln(d) + t),$$

and

$$q(y^*; D) - \mathbb{E}q(y; D) \leq \frac{2\Delta}{\epsilon}(\ln(d) + 1).$$

■

Report Noisy Max with Laplace Noise has essentially the same guarantees as the exponential mechanism (on the same discrete domain), but performs better in practice. It turns out that the exponential mechanism is exactly equivalent to RNM with noise added from a different distribution, the Gumbel distribution with parameter $\beta = \frac{2\Delta}{\epsilon}$.

$$\text{Gumbel}(\beta): \text{ pdf } h_\beta(y) = \frac{1}{\beta}\exp(-\frac{x}{\beta} + e^{-\frac{x}{\beta}}). \tag{4.14}$$

This equivalence is known in the machine learning literature as the Gumbel max trick, where the exponential mechanism is called the Gibbs or softmax distribution. It turns out that RNM with Laplace noise is equivalent to a different process, Permute and Flip.

References