KING ABDULLAH UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Homework 2

Konstantin Burlachenko

Assignment in Contemporary Topics in Computer Security.
Last updated on October 15, 2021

Contents

Task 1: Name and Shame Mechanism

Let's assume we have a pair of neighborhoods D, D' that differs in one item data indexed by i. Now Lets consider two events: Good $= \{Y_i = \text{nothing}\}$, and Bad $= \{Y_i \neq \text{nothing}\}$. This two events is a full group for sample space and so we can use formula of full probability:

$$\begin{aligned}\mathbf{P}(A(D) \in E) = \mathbf{P}(A(D) \in E \cap (\text{Good} \cup \text{Bad}) = \\ \mathbf{P}(A(D) \in E \cap \text{Good}) + \mathbf{P}(A(D) \in E \cap \text{Bad}) = \\ \mathbf{P}(A(D) \in E | y_i = \text{nothing})\mathbf{P}(y_i = \text{nothing}) + \mathbf{P}(A(D) \in E \cap \text{Bad}) = \\ \mathbf{P}(A(D') \in E | y_i = \text{nothing})\mathbf{P}(y_i = \text{nothing}) \cdot \exp(0) + \mathbf{P}(A(D) \in E \cap \text{Bad})\end{aligned}$$

The last equality holds because D, D' is exactly the same in the case of conditioning in $\{y_i = \text{nothing}\}$. Next, we need to bound the first and second terms in the last expression.

First term:

$$\begin{aligned}\mathbf{P}(A(D') \in E | y_i = \text{nothing})\mathbf{P}(y_i = \text{nothing}) \cdot \exp(0) = \\ \mathbf{P}(A(D') \in E \cap \{y_i = \text{nothing}\}) \cdot \exp(0) \leq \\ \mathbf{P}(A(D') \in E) \cdot \exp(0)\end{aligned}$$

Second term:

$$\begin{aligned}\mathbf{P}(A(D) \in E \cap \text{Bad}) = \mathbf{P}(A(D) \in E | A(D) \in \text{Bad})\mathbf{P}(A(D) \in \text{Bad}) = \\ \mathbf{P}(A(D) \in E | A(D) \in \text{Bad}) \cdot \delta \leq 1 \cdot \delta\end{aligned}$$

So we have shown:

$$\mathbf{P}(A(D) \in E) \leq \mathbf{P}(A(D') \in E) + \delta, \forall E, \forall D \sim D'$$

This means that this mechanism is $(0, \delta)$-DP. Remark: some of the steps repeat steps Lemma 6.4 from the Lecture Notes 6.

Task 2:Noisy-max with Laplace Noise

Noisy-Max Mechanism lie in solving privately task with maximizing score function $q(Y, D)$ when input domain Y is finite and that care about privacy aspects in datasetS D. The noisy-max mechanism returns: $\arg\max_y (q(y, D) + Z_y)$ where Z_y is i.i.d. noise.

We consider the case when the output of the Noisy-Max mechanism is value i. Also, we fix the vector of all random variables $Z = \{Z_1, Z_2, \dots\}$ for specific values from $\mathbb{R}$, except position i in which we do not condition. We denote such a random vector with almost all fixed values, except position i as Z_{-i}. Next, we consider deterministic value z^*:

$$z^* := \arg\min_z q(i, D) + z > q(j, D) + Z_j, \forall j \neq i$$

If fix all $z_j, j \neq i$ then event $\{Y = i\}$ is the same as $\{z_i \geq z^*\}$, because that event will give us plausible additive shifts in the Noisy Max mechanism that will give as output $\{Y = i\}$. In that case, $\{z_i \geq z^*\}$ condition is necessary and sufficient.

Via using global Δ - sensitivity $\sup_{y \in Y, D \sim D'} |q(y, D) - q(y, D')| \leq \Delta$ we have firstly:

$$
\begin{aligned}
|q(y, D) - q(y, D')| \leq \Delta &\implies q(y, D') + \Delta \geq q(y, D), \forall y, D \sim D' \\
|q(y, D') - q(y, D)| \leq \Delta &\implies q(y, D') - \Delta \leq q(y, D), \forall y, D \sim D'
\end{aligned}
$$

We use this inequalities to modify condition on z^* workable with neighborhood D':

$$
\begin{gathered}
q(i, D) + z^* > q(j, D) + Z_j, \forall j \neq i \implies \\
(q(i, D') + \Delta) + z^* > (q(j, D') - \Delta) + Z_j, \forall j \neq i \implies \\
q(i, D') + z^* + 2\Delta > q(j, D') + Z_j, \forall j \neq i
\end{gathered}
$$

If $\{z_i > z^* + 2\Delta\}$ then Noisy-max mechanism applied for dataset D' conditioned on event Z_{-i} will give us output i. We can use this fact to measure probability of that two equivalent events: $\mathbf{P}(Y = i | D', Z_{-i}) \geq \mathbf{P}(z_i \geq z^* + 2\Delta)$.

We use the sign $\geq$ sign because $\{z_i \geq z^* + 2\Delta\}$ is only a sufficient condition for the left-hand side event.

Now we use Laplace distribution for additive noise in form where Z_i are independent r.v. with pdf: $h(y, \lambda = \Delta/\varepsilon) = \dfrac{1}{2\lambda} \exp(-|y|/\lambda) = \dfrac{\varepsilon}{2\Delta} \exp(-\frac{\varepsilon}{\Delta}|y|)$ with $dom(h) = \mathbb{R}$.

$$
\begin{aligned}
\mathbf{P}(Y = i | D', Z_{-i}) \geq \mathbf{P}(z_i \geq z^* + 2\Delta) = \int_{z^*+2\Delta}^{+\infty} \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon}{\Delta}|y|\right) dy, |y = t + 2\Delta|, \\
\int_{z^*}^{+\infty} \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon}{\Delta}|t + 2\Delta|\right) dt \geq \int_{z^*}^{+\infty} \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon}{\Delta}(|t| + |2\Delta)|\right) = \\
\exp(-\varepsilon) \int_{z^*}^{+\infty} \frac{\varepsilon}{2\Delta} \exp\left(-\frac{\varepsilon}{\Delta}(|t|)\right) = \exp(-\varepsilon)\mathbf{P}(Z_i > z^*) = \exp(-\varepsilon)\mathbf{P}(Y = i | D, Z_{-i})
\end{aligned}
$$

In the middle of derivation we have used tringale inequality $|a+b| \leq |a|+|b|$ in combination with using it with using it for decreasing function $\exp(-x)$. After multiply by $\exp(\varepsilon)$ We have proved:

$$
\mathbf{P}(Y = i | D, Z_{-i}) \leq \mathbf{P}(Y = i | D', Z_{-i}) \exp(\varepsilon)
$$

Next $\forall D \sim D'$ and if denoted $\mathcal{Z}$ as a range of Z_{-i} we can margianlized out dependence on Z_{-i} and apply inequlity for arbitarily event E:

$$
\begin{aligned}
\mathbf{P}(Y = i | D', Z_{-i}) \leq \exp(\varepsilon)\mathbf{P}(Y = i | D, Z_{-i}) \implies \\
\mathbf{P}(A(D') = i, Z_{-i})\mathbf{P}(Z_{-i}) \leq \exp(\varepsilon)\mathbf{P}(A(D) = i, Z_{-i})\mathbf{P}(Z_{-i}) \implies \\
\mathbf{P}(A(D') = i, Z_{-i}) \leq \exp(\varepsilon)\mathbf{P}(A(D) = i, Z_{-i}) \implies \text{(We marginalized out)} \\
\int_{Z_{-i} \in \mathcal{Z}} \mathbf{P}(A(D') = i, Z_{-i}) d(Z_{-i}) \leq \int_{Z_{-i} \in \mathcal{Z}} \exp(-\varepsilon)\mathbf{P}(A(D) = i, Z_{-i}) d(Z_{-i}) \implies \\
\mathbf{P}(A(D') = i) \leq \exp(\varepsilon)\mathbf{P}(A(D) = i) \implies \\
\mathbf{P}(A(D') = k) \leq \exp(\varepsilon)\mathbf{P}(A(D) = k), \forall k \in E \implies \sum_{k \in E} \mathbf{P}(A(D') = k) \leq \sum_{k \in E} \exp(\varepsilon)\mathbf{P}(A(D) = k) \implies \\
\mathbf{P}(A(D') \in E) \leq \exp(\varepsilon)\mathbf{P}(A(D) \in E) \blacksquare
\end{aligned}
$$

Task 3: Adding Uniform Noise to count query

We have a count query defined as $f : \{0,1\}^n \to \mathbf{R}$ with $f(D) = \sum_{i=1}^n x_i$ and we release $A(D) = f(D) + Z$ where $Z \sim U_{[-\lambda,+\lambda]}$. Z is real valued r.v. which has p.d.f $f(z) = \frac{1}{2\lambda} \cdot \mathbf{1}\{z \in [-\lambda,+\lambda]\}$.

Global sensitivity is $\Delta = 1$, because in case of changing single data point $f(D) - f(D') \in -1, 1 \implies |f(D) - f(D')| \leq 1$.

If W has p.d.f $p_w(w)$, then $W + c, \forall c \in \mathbb{R}$ has p.d.f $p_w(w - c)$:

$$\mathbf{P}(c + W < t) = \mathbf{P}(W < t - c) = \int_{-\infty}^{t-c} p_w(w)dw = |u - c = w| = \int_{-\infty}^{t} p_w(u - c)du.$$

This proves that $c + W$ has p.d.f. $p_w(u - c)$. Now let's take a look into privacy loss:

$$l_{D,D'}(y) = \ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)} = \ln \frac{U_{[-\lambda,+\lambda]}(y - f(D))}{U_{[-\lambda,+\lambda]}(y - f(D'))} = \ln \frac{\frac{1}{2\lambda}\mathbf{1}\{y - f(D) \in [-\lambda, \lambda]\}}{\frac{1}{2\lambda}\mathbf{1}\{y - f(D') \in [-\lambda, \lambda]\}} =$$
$$\ln \frac{\mathbf{1}\{y - f(D) \in [-\lambda, \lambda]\}}{\mathbf{1}\{y - f(D') \in [-\lambda, \lambda]\}} = ln \frac{\mathbf{1}\{y + (f(D') - f(D)) \in [f(D') - \lambda, f(D') + \lambda]\}}{\mathbf{1}\{y \in [f(D') - \lambda, f(D') + \lambda]\}}$$

First remark is when both numerator and denominator has zero value, we can assume that $\frac{0}{0} = 1$. This is so because in fact we're interesting to bound $(0 = p_{A(D)}(y)) \leq \exp(0) \cdot (0 = p_{A(D')}(y))$. To bound $0 \leq 0$ we can use arbitrarily finite multiple, but in particular we can use $\exp(0)$.

Now we consider *Good* case when for expression $\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)}$:

1. Numerator and denominator both attains values 1, 1 respectively.
2. Numerator and denominator both attains values 0, 0 respectively.
3. Numerator and denominator both attains values 0, 1 respectively.

For *Good* event we can bound for all that case $\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)} \leq 0$ For *Good* event this mechanism $\varepsilon = 0$ DP.

Now we consider *Bad* case when for expression $\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)}$:

1. **Numerator and denominator both attains values** $1, 0$ **respectively.**

For Bad event we will have we will have $\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)} > 0$. Due to Theorem 6.14, Lecture 6 – it's enough to demonstrate that this mechanism is $(0, \delta)$-DP it's enough to show that $\mathbf{P}[\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)} > 0] \leq \sigma$. As alternative we can also prove that probability of the *Bad* event is bound by σ, which is by Lemma 6.4 will be also enough to prove that this mechanism is $(0, \sigma)$ DP.

$$\mathbf{P}[\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)} > 0] =$$

$$\mathbf{P}[\{y + (f(D') - f(D)) \in [f(D') - \lambda, f(D') + \lambda]\} \cap \{y < f(D') - \lambda \cup y > f(D') + \lambda\}] =$$

$$\mathbf{P}[\{y + (f(D') - f(D)) \in [f(D') - \lambda, f(D') + \lambda]\} \cap \{y < f(D') - \lambda\}] +$$

$$\mathbf{P}[\{y + (f(D') - f(D)) \in [f(D') - \lambda, f(D') + \lambda]\} \cap \{y > f(D') + \lambda\}]$$

We know that $-\Delta \leq f(D') - f(D) \leq \Delta$, and we will use this knowledge to bound last expression, also via taking into account sign of $f(D) - f(D')$ that is fixed, but which we don't know:

$$\mathbf{P}[\ln \frac{p_{A(D)}(y)}{p_{A(D')}(y)} > 0] \leq$$

$$\mathbf{P}[\{y \in [f(D') - \lambda - \Delta, f(D') - \lambda]\}] \cdot \mathbf{1}\{f(D') - f(D) \geq 0\} +$$

$$0 \cdot \mathbf{1}\{f(D') - f(D) < 0\} +$$

$$0 \cdot \mathbf{1}\{f(D') - f(D) > 0\} +$$

$$\mathbf{P}[\{y \in [f(D') + \lambda, f(D') + \lambda + \Delta]\}] \cdot \mathbf{1}\{f(D') - f(D) \leq 0\} =$$

$$(\Delta)/(2\lambda) \cdot \mathbf{1}\{f(D') - f(D) \geq 0\} + (\Delta)/(2\lambda) \cdot \mathbf{1}\{f(D') - f(D) < 0\} = \frac{\Delta}{2\lambda} \leq \sigma.$$

The second term in summation is zero because if $f(D') - f(D) < 0$:
$\{y < f(D') - \lambda\} \cap \{y + (f(D') - f(D)) > f(D') - \lambda\} = 0$.

The third term in summation is zero because if $f(D') - f(D) > 0$
$\{y > f(D') + \lambda\} \cap \{y + (f(D') - f(D)) \leq f(D') + \lambda\} = 0$.

To achieve $(0, \sigma)$ - DP we need to have:

$$\lambda \geq \frac{\Delta}{2\sigma}$$

In our counting query $\Delta = 1$ and so the final bound for parameter of the additive uniform noise for mechanism is:

$$\lambda \geq \frac{1}{2\sigma}.$$

Do both ε and σ matter in setting? If we fix the nature of noise, then for set noise, we have a single parameter of the noise λ. The set of this parameter does not affect the result value ε. It's always 0. From another point of view, it's impossible to make this mechanism pure-DP for any finite value of λ. Intuitively with small probabilities, the **Bad** situation may happen in which the information loss is infinity huge.

When $\sigma < \frac{1}{n}$, will this mechanism produce useful information? Global sensitivity is equal to 1 that does not depend on n. To have $(0, \sigma)$ guarantee really we need only to have $\sigma > \frac{1}{2\lambda}$. So in case $1/n \geq 1/(2\lambda) \iff \lambda \geq n/2$ we can preserve $(0, \sigma)$ - DP.

Next, because: $A(D) = f(D) + Z$ where $Z \sim U_{[-\lambda, +\lambda]}$ we know that $E[A(D)] = f(D)$. Also $Var[A(D)] = Var[f(D) + Z] = Var[Z] = \frac{1}{12}(\lambda - (-\lambda))^2 = \frac{1}{12}4\lambda^2 = \lambda^2/3$. Chebychev's inequality can be formalized in that form: $\mathbf{P}(|X = E[X]| > k\sigma) \leq \frac{1}{k^2}$.

By Chebychev inequality with reasonable probability we can say that $|A(D) - f(D)| \leq \mathbf{O}(\lambda)$. Constant in O-notation for reasonable probability maybe 10. To make λ as small as possible we can select $\lambda = n/2$ and we can make conclusion: $|A(D) - f(D)| \leq \mathbf{O}(n)$.

Unfortunately, this is **not useful information for compute numerical values for counting query**, because we know from the construction of counting query that $f(D)$ a $f(D) \in \{0, 1, \dots, n\}$.

Example of useful information can be obtained with using: $\sigma = \dfrac{1}{n^\alpha}$, for example we can select $\lambda = n^\alpha/2$ to preserve $(0, \sigma)$ - DP and decrease interval.

And with high probability: $|A(D) - f(D)| \leq \mathbf{O}(\lambda) = \mathbf{O}(n^\alpha), \forall^\alpha \in (0, 1)$.

Task 4: Implementation of Noisy-max Mechanism and Exponential Mechanism

The exponential and noisy-max mechanism allows us to privately select an object with a score comparable to the best.

The implementation. Implementation has been done in Python with using Numpy library as a backend for matrix-vector operations and random number generations. The matplotlib library has been used to plot graphics. The Global Sensitivity parameter for a selection problem $\Delta = \sup_{Y \in \mathcal{Y}} \sup_{D \sim D'} |q(y, D) - q(y, D')|$ has been found via brute-force(complete discrete search) for a given problem with using parallelization across CPU threads. The source code is locating in *experiment_noisy_max.py*.

Problem setup. There are $n = 8$ clients which has prices for assets that they want to buy, the prices of the clients lie in the set $X = \{0.01, 0.10, 0.20, 0.30, 0.50\}$. The private dataset $D \in X^n$ contains prices $D = (X_1, X_2, X_3, X_4, X_5, X_5, X_3, X_1)$. The set of prices is finite and it consist of prices $Y = \{0.08, 0.12, 0.25, 0.35, 0.45, 0.50\}$.

The result os experiments during apply max selection mechanisms for different values of $\varepsilon \in [0.001, 10]$ are presented in Figure 1. The error bars demonstrate error bars for estimated standart deviation across 900 computed experiments.

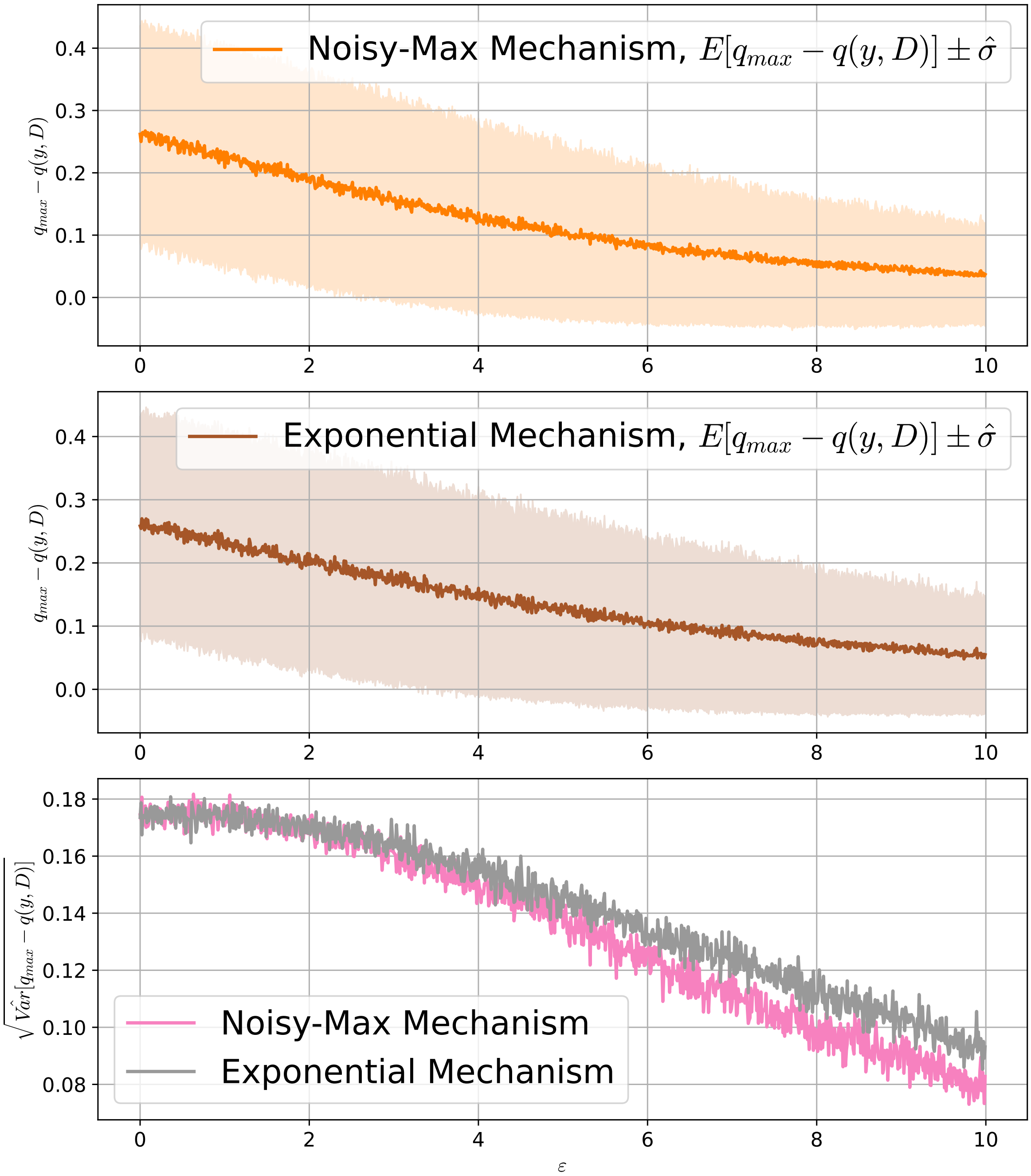

Figure 1: Dependency of errors in DP mechanism as a function of ε. First two plot contains error of for computing $f(x)$ with error bars.

Findings.

1. The most computational demand part is finding the Global Sensitivity parameter because cardinality of X^n is 390625 and this is only all possible unique $D \in X^n$. For each dataset D, there are near $n \cdot |Y| = 8 \cdot 6 = 48$ neighborhood dataset D', and finally, we need to perform this analysis for each Y and in our case $|Y| = 6$. So, finally to compute global sensitivity we need to perform 112 500 000 calls for query function evaluation $q(y, D)$. So having the ability to compute fast global sensitivity is important to consider more big experiments.

2. The asymptotic behavior of the two algorithms is the same. We see that as a bigger violation of privacy we allow via setup bigger ε the more precise algorithms start to be in terms of providing an answer that is more near to $q_{max} = \max_{y \in \mathcal{Y}} q(y, D)$.

3. The asymptotic dependence from plots for both algorithms has the following characteristics $q_{max} - \mathbf{E}[q(y, D)] \propto 2\Delta/\varepsilon$ and this coincides with Theory for utility for both of this two algorithms (Lecture 5, Theorem 0.7, 0.10).

4. Finally, we can observe that the behavior of Noisy-Max is slightly better. It produces a solution with less variance, and also in computer systems, compute exp is unstable, cost. So from the practical point of view, Noisy-Max can be a bit more superior method.

5. Theoretical bounds are not worst, but they describe bounds achievable with high probability. As we see from that instance of the problem, the Theory predicts the behavior of algorithms very well.

6. Big values of ε introduce less noise into mechanism and as consequence there r.v. $q_{max} - q(y, D)$ has less variance as ε increases,