CSE711: Topics in Differential Privacy **Spring 2016**

Lecture 1 Introduction to Differential Privacy: January 28

Lecturer: Marco Gaboardi *Scribe: Marco Gaboardi*

This note introduces differential privacy [DMNS06], one important mechanism (the Laplace mechanism) that can be used to guarantee it, and some important properties that follow from the definition. We will jump directly in the technical definitions and we will leave the discussion about why this definition makes sense to the end of the class. This note is a compressed summary of Chapters 1-3 of [DR14].

1.1 Randomized Algorithms

Differential privacy is a property of some randomized algorithms. So, we will start by defining what a *randomized algorithm* is. We will mostly focus on randomized algorithms whose probabilitistic space is discrete. To formalize this intuition we will use the notion of probability simplex.

Definition 1.1 (Probability simplex) *Given a discrete set B, the* probability simplex *over B, denoted $\Delta(B)$ is the set:*

$$\Delta(B) = \left\{ x \in \mathbb{R}^{|B|} : \forall i, x_i \geq 0, \text{ and } \sum_{i=1}^{|B|} x_i = 1 \right\}$$

Now we can describe randomized algorithms.

Definition 1.2 (Randomized Algorithms) *A* randomized algorithm *$\mathcal{M}$ with domain A and range B is an algorithm associated with a total map $M : A \to \Delta(B)$. On input $a \in A$, the algorithm $\mathcal{M}$ outputs $\mathcal{M}(a) = b$ with probability $(M(a))_b$ for each $b \in B$. The probability space is over the coin flips of the algorithm $\mathcal{M}$.*

1.2 Differential Privacy

To define differential privacy we need first to define what we want to protect, i.e. what is the formal datum that we aim to protect. In the differential privacy literature this corresponds often to the *presence or absence of an individual* in a database. Differential privacy aims at sanitizing a data analysis so that the presence or absence of an individual in a study cannot be distinguished by just observing the result of the analysis.

For simplicity we will avoid implementation details and we will consider databases as *histograms*. Given a universe $\mathcal{X}$ an histogram over $\mathcal{X}$ is an object in $\mathbb{N}^{|\mathcal{X}|}$. We can bake in the *presence or absence of an individual* notion in a definition of *distance* between databases.

Definition 1.3 (Distance Between Databases) *The ℓ_1 norm of a database $x \in \mathbb{N}^{|\mathcal{X}|}$ is denoted $\|x\|_1$ and is defined as:*

$$\|x\|_1 = \sum_{i=1}^{|\mathcal{X}|} x_i$$

The ℓ_1 distance between two databases x and y is defined as $\|x - y\|_1$.

We will call two databases $x, y \in \mathbb{N}^{\mathcal{X}}$ *adjacent* if $\|x - y\|_1 \leq 1$. Notice that two databases are adjacent if they are equal or if they differ for the presence or absence of a single individual.

We can now define differential privacy.

Definition 1.4 (Differential privacy [DMNS06]) *A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ and range R: is (ϵ, δ)-differentially private for $\epsilon, \delta \geq 0$ if for every adjacent $x, y \in \mathbb{N}^{|\mathcal{X}|}$ and for any subset $S \subseteq R$ we have*

$$\Pr[\mathcal{M}(x) \in S] \leq \exp(\epsilon) \Pr[\mathcal{M}(y) \in S] + \delta. \tag{1.1}$$

Notice that the definition of differential privacy depends on two parameter ϵ and δ. We will call them the *privacy parameters.*

Often we will consider the simpler case where $\delta = 0$, in this case notice that we can rewrite the requirement of differential privacy as requiring that for every adjacent $x, y \in \mathbb{N}^{|\mathcal{X}|}$ and for any $r \in R$ we have

$$\exp(-\epsilon) \leq \frac{\Pr[\mathcal{M}(x) = r]}{\Pr[\mathcal{M}(y) = r]} \leq \exp(\epsilon)$$

The quantity

$$\ln\Big(\frac{\Pr[\mathcal{M}(x) = r]}{\Pr[\mathcal{M}(y) = r]}\Big)$$

is often called the *privacy loss* of the algorithm $\mathcal{M}$.

1.3 Randomized response

There are several ways for designing differentially private algorithms. We will see some of them in the rest of the semester. Here we will consider a simple first example of a differentially private algorithm:

```
RandResponse(x)
begin
flip a coin
    if head then
        if embarassing_question(x) then
            answer yes
        else
            answer no
        endif
    else
        flip a coin
        if head then
            answer yes
        else
            answer no
        endif
    endif
end
```

The algorithm above is an instance of a more general class of algorithms commonly referred to as *randomize response*. We can easily prove that the instance above of randomized response is differentially private.

Claim 1.5 (Privacy for Randomized Response) *The algorithm* RandResponse *is* $(\ln 3, 0)$*-differentially private.*

Proof: We can reason in a similar way for each possible answer, so let us focus on the answer *yes*. Let us consider two neighbor databases x and y. In particular, let us assume that $\mathsf{embarassing\_question}(x)$=*yes* and $\mathsf{embarassing\_question}(y)$=*no*. Then, a case analysis shows that for every z

$$\Pr[\text{Response}{=}\textit{yes} \mid \mathsf{embarassing\_question}(z) = \textit{yes}] = 3/4.$$

Specifically, when $\mathsf{embarassing\_question}$(z)=*yes* the outcome is *yes* if the first coin comes up tails (probability 1/2) or the first and second come up heads (probability 1/4)). Similarly,

$$\Pr[\text{Response}{=}\textit{yes} \mid \mathsf{embarassing\_question}(z) = \textit{no}] = 1/4$$

(first comes up heads and second comes up tails; probability 1/4). We can also apply a similar reasoning to the case of a *no* answer. Putting these together and instantiating them on x and y we obtain:

$$\frac{\Pr[\text{Response}{=}\textit{yes} \mid \mathsf{embarassing\_question}(x) = \textit{yes}]}{\Pr[\text{Response}{=}\textit{yes} \mid \mathsf{embarassing\_question}(y) = \textit{no}]} = \frac{3/4}{1/4} = \frac{\Pr[\text{Response}{=}\textit{no} \mid \mathsf{embarassing\_question}(x) = \textit{no}]}{\Pr[\text{Response}{=}\textit{no} \mid \mathsf{embarassing\_question}(y) = \textit{yes}]} = 3$$

■

In general, when designing a differentially private algorithm we are also interested in having a formal result stating how good is the answer that the algorithm gives us. This is usually formulated as an *accuracy* result that can assume different forms. In the case of randomized response we have the following result.

Claim 1.6 (Accuracy for Randomized response) *Let* $r = \mathsf{RandResponse}(x)$*. Then*

$$\Pr\Big[\mathsf{embarassing\_question}(x) = \mathsf{RandResponse}(x)\Big] = 3/4$$

1.4 The Laplace mechanism

We will now see another simple method to obtain differential privacy for numeric queries/functions. This method adds to the result of a numeric query some statistical noise distributed accordingly to the Laplace distribution.

Definition 1.7 (Laplace Distribution) *The Laplace distribution (centered at 0) with scale b is the probability distribution with probability density function:*

$$\mathsf{Lap}(x|b) = \frac{1}{2b}\exp\Big(-\frac{|x|}{b}\Big)$$

The variance of this distribution is $\sigma^2 = 2b^2$

The noise that we want to add is described by the parameter b. To ensure a bound on the privacy loss this has to be calibrated to the possible influence that a single individual can have on the result of the numeric query. This influence is captured by the notion of *global sensitivity*.

Definition 1.8 (ℓ_1 global sensitivity) *The ℓ_1* global sensitivity *of a function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$ is:*

$$\Delta f = \max \left\{ \|f(x) - f(y)\|_1 \;\middle|\; x, y \in \mathbb{N}^{|\mathcal{X}|} \text{ adjacent} \right\}$$

In general sensitivity can be defined also for norms other than ℓ_1.

We can now define the Laplace mechanism.

Definition 1.9 (Laplace Mechanism) *Given any function $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$, the Laplace mechanism is defined as:*

$$\mathcal{M}_L(x, f(\cdot), \epsilon) = f(x) + Y$$

where Y is a random variable drawn from $\mathsf{Lap}(y|\frac{\Delta f}{\epsilon})$.

We want to prove two properties of the Laplace mechanism: that it ensures differential privacy and that it has a non-trivial accuracy. Let's start by proving that it ensures differential privacy.

Theorem 1.10 (Privacy of the Laplace mechanism) *The Laplace mechanism ensures $(\epsilon, 0)$-differential privacy.*

Proof: Consider $x, y \in \mathbb{N}^{|\mathcal{X}|}$, $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$, and let p_x and p_y denote the probability density function of $\mathcal{M}_L(x, f(\cdot), \epsilon)$ and $\mathcal{M}_L(y, f(\cdot), \epsilon)$, respectively. We compare them at an arbitrary point $z \in \mathbb{R}$. We have:

$$\begin{aligned}
\frac{p_x(z)}{p_y(z)} &= \frac{\exp\left(-\frac{\epsilon|f(x)-z|}{\Delta f}\right)}{\exp\left(-\frac{\epsilon|f(y)-z|}{\Delta f}\right)} \\
&= \exp\left(\frac{\epsilon(|f(y)-z| - |f(x)-z|)}{\Delta f}\right) \\
&\le \exp\left(\frac{\epsilon(|f(y)-f(x)|)}{\Delta f}\right) \\
&= \exp\left(\frac{\epsilon(\|f(y)-f(x)\|_1)}{\Delta f}\right) \\
&\le \exp(\epsilon)
\end{aligned}$$

■

We also can prove that it has a non trivial accuracy.

Theorem 1.11 (Accuracy of the Laplace mechanism) *Let $f : \mathbb{N}^{|\mathcal{X}|} \to \mathbb{R}$, and let $r = \mathcal{M}_L(x, f(\cdot), \epsilon)$. Then $\forall p \in (0, 1]$:*

$$\Pr\left[|f(x) - r| \ge \left(\frac{\Delta f}{\epsilon}\right) \ln\left(\frac{1}{p}\right)\right] = p$$

Proof: By definition of the Laplace mechanism we have:

$$\Pr\left[|f(x) - r| \ge \left(\frac{\Delta f}{\epsilon}\right) \ln\left(\frac{1}{p}\right)\right] = \Pr\left[|Y| \ge \left(\frac{\Delta f}{\epsilon}\right) \ln\left(\frac{1}{p}\right)\right]$$

where Y is drawn from $\mathsf{Lap}\big(y|\frac{\Delta f}{\epsilon}\big)$. The Laplace distribution has a tail bound that guarantees that if Z is drawn from $\mathsf{Lap}\big(y|b\big)$ then:

$$\Pr\left[|Z| \ge b\,t\right] = \exp(-t)$$

Applying this fact we have

$$\Pr\Big[|Y| \geq \Big(\frac{\Delta f}{\epsilon}\Big)\ln\Big(\frac{1}{p}\Big)\Big] = \exp\Big(-\ln\Big(\frac{1}{p}\Big)\Big) = p$$

■

The Laplace mechanism ensures $(\epsilon, 0)$-differential privacy. There is a similar mechanism sampling from the gaussian distribution that is useful to ensure (ϵ, δ)-differential privacy.

1.5 Some properties

We conclude this note by observing three important properties of differential privacy. These follows directly from the definition and they are fundamental facts that gave to differential privacy strong credit.

1.5.1 Post-processing

The first property ensures that the results of differentially private computations can be safely released because any post-processing computation will also be differentially private.

Proposition 1.12 (Post-processing) *Let $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to R$ be a randomized algorithm that is (ϵ, δ)-differentially private. Let $f : R \to R'$ be an arbitrary deterministic mapping. Then $f \circ \mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to R'$ is (ϵ, δ) differentially private.*

Proof: Fix any pair of neighboring databases x, y with $\|x - y\|_1 \leq 1$, and fix any event $S \subseteq R'$. Let $T = \{r \in R : f(r) \in S\}$. We have

$$\begin{aligned} \Pr[f(\mathcal{M}(x)) \in S] &= \Pr[\mathcal{M}(x) \in T] \\ &\leq \exp(\epsilon) Pr[\mathcal{M}(y) \in T] + \delta \\ &= \exp(\epsilon) Pr[f(\mathcal{M}(y)) \in S] + \delta \end{aligned}$$

■

This result can also be generalized to arbitrary randomized mappings, see [DR14].

1.5.2 Group Privacy

The second property illustrates how differential privacy can be used to protect also the privacy of groups rather than single individuals.

Proposition 1.13 (Group Privacy) *Let $\mathcal{M} : \mathbb{N}^{|\mathcal{X}|} \to R$ be a randomized algorithm that is $(\epsilon, 0)$-differentially private. Then, $\mathcal{M}$ is $(k\epsilon, 0)$-differentially private for groups of size k. That is, for databases $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq k$ and for all $S \subseteq R$ we have*

$$\Pr[\mathcal{M}(x) \in S] \leq \exp(k\epsilon) \Pr[\mathcal{M}(y) \in S]$$

Proof: Fix any pair of databases x, y with $\|x - y\|_1 \leq k$. Then, we have databases $z_0, z_1, \ldots, z_k$ such that $z_0 = x$, $z_k = y$ and $\|z_i - z_{i+1}\|_i \leq 1$. Fix also any event $S \subseteq R'$. Then, we have have

$$\begin{aligned}
\Pr[\mathcal{M}(x) \in S] &= \Pr[\mathcal{M}(z_0) \in S] \\
&\leq \exp(\epsilon) \Pr[\mathcal{M}(z_1) \in S] \\
&\leq \exp(\epsilon)(\exp(\epsilon) \Pr[\mathcal{M}(z_2) \in S]) = \exp(2\epsilon) \Pr[\mathcal{M}(z_2) \in S] \\
&\leq \cdots \\
&\leq \exp(k\epsilon) \Pr[\mathcal{M}(z_k) \in S] = \exp(k\epsilon) \Pr[\mathcal{M}(y) \in S]
\end{aligned}$$

■

1.5.3 Composition

Finally, the third property ensures that we can safely compose differentially private computations with a controlled degradation of the privacy loss.

Proposition 1.14 (Standard composition for $(\epsilon, 0)$-differential privacy) *Let $\mathcal{M}_1 : \mathbb{N}^{|\mathcal{X}|} \to R_1$ be an $(\epsilon_1, 0)$-differentially private algorithm and let $\mathcal{M}_2 : \mathbb{N}^{|\mathcal{X}|} \to R_2$ be an $(\epsilon_2, 0)$-differentially private algorithm. Then their composition defined to be $\mathcal{M}_{1,2} : \mathbb{N}^{|\mathcal{X}|} \to R_1 \times R_2$ by the mapping $\mathcal{M}_{1,2}(x) = (\mathcal{M}_1(x), \mathcal{M}_2(x))$ is $(\epsilon_1 + \epsilon_2, 0)$-differentially private.*

Proof: Fix any pair of databases x, y with $\|x - y\|_1 \leq k$. Fix also a pair of output $(r_1, r_2) \in R_1 \times R_2$. We have:

$$\begin{aligned}
\frac{\Pr[\mathcal{M}_{1,2}(x) = (r_1, r_2)]}{\Pr[\mathcal{M}_{1,2}(y) = (r_1, r_2)]} &= \frac{(\Pr[\mathcal{M}_1(x), \mathcal{M}_2(x)) = (r_1, r_2)]}{(\Pr[\mathcal{M}_1(y), \mathcal{M}_2(y)) = (r_1, r_2)]} \\
&= \frac{\Pr[\mathcal{M}_1(x) = r_1] \Pr[\mathcal{M}_2(x) = r_2]}{\Pr[\mathcal{M}_1(y) = r_1] \Pr[\mathcal{M}_2(y) = r_2]} \\
&= \Big(\frac{\Pr[\mathcal{M}_1(x) = r_1]}{\Pr[\mathcal{M}_1(y) = r_1]}\Big)\Big(\frac{\Pr[\mathcal{M}_2(x) = r_2]}{\Pr[\mathcal{M}_2(y) = r_2]}\Big) \\
&\leq \exp(\epsilon_1) \exp(\epsilon_2) = \exp(\epsilon_1 + \epsilon_2).
\end{aligned}$$

Similarly, we can prove $\frac{\Pr[\mathcal{M}_{1,2}(x) = (r_1, r_2)]}{\Pr[\mathcal{M}_{1,2}(y) = (r_1, r_2)]} \geq \exp(-(\epsilon_1 + \epsilon_2))$. ■

Differential privacy supports other notions of composition for different possible degradations of the privacy parameters, see [DR14].

References

[DMNS06] Cynthia Dwork and Frank McSherry and Kobbi Nissim and Adam Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," *Proceedings of the Theory of Cryptography Conference, TCC*, 2006, pp. 265–284.

[DR14] Cynthia Dwork and Aaron Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends in Theoretical Computer Science*, Vol 9, Nos 3–4, pp. 211–407, 2014.