CSE660
Differential Privacy

September 20, 2017

Marco Gaboardi
Room: 338-B
gaboardi@buffalo.edu
http://www.buffalo.edu/~gaboardi

(ε,δ)-Differential Privacy

Definition

Given $\varepsilon, \delta \geq 0$, a probabilistic query $Q: X^n \rightarrow R$ is (ε,δ)-differentially private iff

for all adjacent database b_1, b_2 and for every $S \subseteq R$:

$$\Pr[Q(b_1) \in S] \leq \exp(\varepsilon)\Pr[Q(b_2) \in S] + \delta$$

Randomized Response

Algorithm 1 Pseudo-code for Randomized Response

1: **function** RANDOMIZEDRESPONSE(D, q, ϵ)
2: **for** $k \leftarrow 1$ to $|D|$ **do**
3: $$S_i \leftarrow \begin{cases} q(d_i) & \text{with probability } \frac{e^\epsilon}{1+e^\epsilon} \\ \neg q(d_i) & \text{with probability } \frac{1}{1+e^\epsilon} \end{cases}$$
4: **end for**
5: **return** $\frac{(\textbf{sum } S)}{|D|}$
6: **end function**

Randomized Response

Privacy Theorem:
Randomized response is **ε**-differentially private.

Accuracy Theorem:

$$\Pr_{r \leftarrow RR(D,q,\epsilon)}\left[\left|\frac{1+e^{\epsilon}}{e^{\epsilon}-1}\left(r-\frac{1}{1+e^{\epsilon}}\right)-q(D)\right| \geq \frac{1+e^{\epsilon}}{(e^{\epsilon}-1)}\sqrt{\frac{\log(2/\beta)}{2n}}\right] \leq \beta$$

Laplace Mechanism

Algorithm 2 Pseudo-code for the Laplace Mechanism

1: **function** LAPMECH(D, q, ϵ)
2: $\quad Y \xleftarrow{\$} \mathsf{Lap}(\frac{\Delta q}{\epsilon})(0)$
3: $\quad$ **return** $q(D) + Y$
4: **end function**

Laplace Mechanism

Theorem (Privacy of the Laplace Mechanism)
The Laplace mechanism is ε-differentially private.

Accuracy Theorem: $\text{let } r = \mathsf{LapMech}(D, q, \epsilon)$

$$\Pr\left[|q(D) - r| \geq \left(\frac{\Delta q}{\epsilon}\right) \ln\left(\frac{1}{\beta}\right)\right] = \beta$$

Randomized Response vs Laplace

Accuracy for Randomize response: with high probability we have:

$$\left|r - q(D)\right| \leq O\left(\frac{1}{\sqrt{n}}\right)$$

Accuracy for Laplace: with high probability we have:

$$\left|q(D) - r\right| \leq O\left(\frac{1}{n}\right)$$

Some important properties

- Resilience to post-processing
- Group privacy
- Composition

Resilience to Post-processing

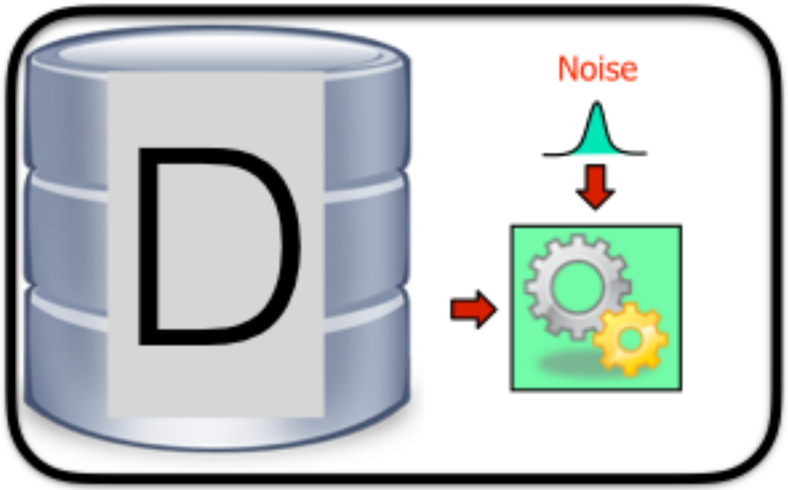
Proposition 1.1 (Post-processing). Let $\mathcal{M} : \mathcal{X}^n \to R$ be a randomized algorithm that is ϵ-differentially private. Let $f : R \to R'$ be an arbitrary deterministic mapping. Then $f \circ \mathcal{M} : \mathcal{X}^n \to R'$ is also ϵ-differentially private.

Group Privacy

Proposition 1.2 (Group Privacy). Let $\mathcal{M} : \mathcal{X}^n \to R$ be a randomized algorithm that is ϵ-differentially private. Then, $\mathcal{M}$ is $k\epsilon$-differentially private for groups of size k. That is, for datasets $D, D' \in \mathcal{X}^n$ such that $D \Delta D' \leq k$ and for all $S \subseteq R$ we have

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

Composition



M_1 is ε_1-DP

M_2 is ε_2-DP

…

M_n is ε_n-DP

The overall process is $(\varepsilon_1+\varepsilon_2+\ldots+\varepsilon_n)$-DP

Composition

Theorem 1.7 (Standard composition for ϵ-differential privacy)**.** Let $\mathcal{M}_1 : \mathcal{X}^n \to R_1$ be an ϵ_1-differentially private algorithm and let $\mathcal{M}_2 : \mathcal{X}^n \to R_2$ be an ϵ_2-differentially private algorithm. Then their composition defined to be $\mathcal{M}_{1,2} : \mathcal{X}^n \to R_1 \times R_2$ by the mapping $\mathcal{M}_{1,2}(D) = (\mathcal{M}_1(D), \mathcal{M}_2(D))$ is $(\epsilon_1 + \epsilon_2)$-differentially private.

Proof. Fix any pair of adjacent datasets $D \sim_1 D'$. Fix also a pair of output $(r_1, r_2) \in R_1 \times R_2$. We have:

$$\frac{\Pr[\mathcal{M}_{1,2}(D) = (r_1, r_2)]}{\Pr[\mathcal{M}_{1,2}(D') = (r_1, r_2)]} = \frac{(\Pr[\mathcal{M}_1(D), \mathcal{M}_2(D)) = (r_1, r_2)]}{(\Pr[\mathcal{M}_1(D'), \mathcal{M}_2(D')) = (r_1, r_2)]}$$

$$= \frac{\Pr[\mathcal{M}_1(D) = r_1] \Pr[\mathcal{M}_2(D) = r_2]}{\Pr[\mathcal{M}_1(D') = r_1] \Pr[\mathcal{M}_2(D') = r_2]} = \Big(\frac{\Pr[\mathcal{M}_1(D) = r_1]}{\Pr[\mathcal{M}_1(D') = r_1]}\Big)\Big(\frac{\Pr[\mathcal{M}_2(D) = r_2]}{\Pr[\mathcal{M}_2(D') = r_2]}\Big)$$

$$\leq \exp(\epsilon_1) \exp(\epsilon_2) = \exp(\epsilon_1 + \epsilon_2).$$

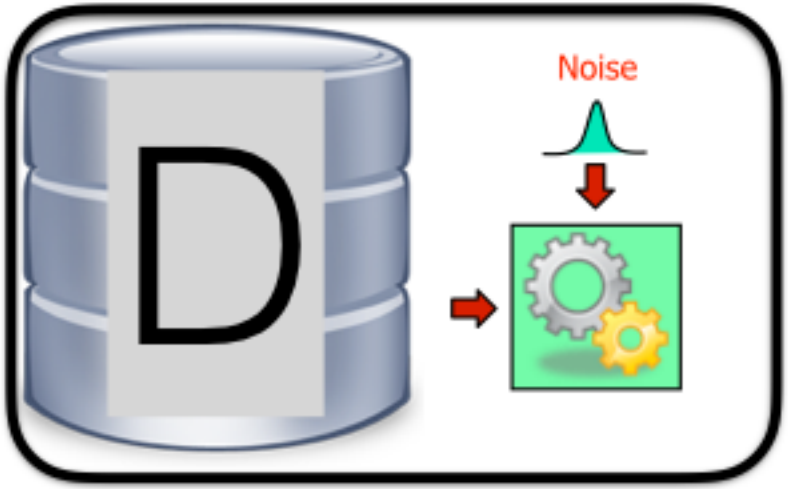
Composition

Question: Why composition is important?

Answer: Because it allows to reason about privacy as a budget!

Composition

Budget=ε_{global} - ε_1 - ε_2 … - ε_n



M_1 is ε_1-DP

M_2 is ε_2-DP

…

M_n is ε_n-DP

Example 1

Let's consider an arbitrary ordered universe domain $\mathcal{X}$ and let's consider the following predicate for $y \in \mathcal{X}$

$$q_y(x) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{otherwise} \end{cases}$$

we call a threshold function the associated counting query

$$q_y : \mathcal{X}^n \to [0, 1]$$

Question: What is the sensitivity?

Example I

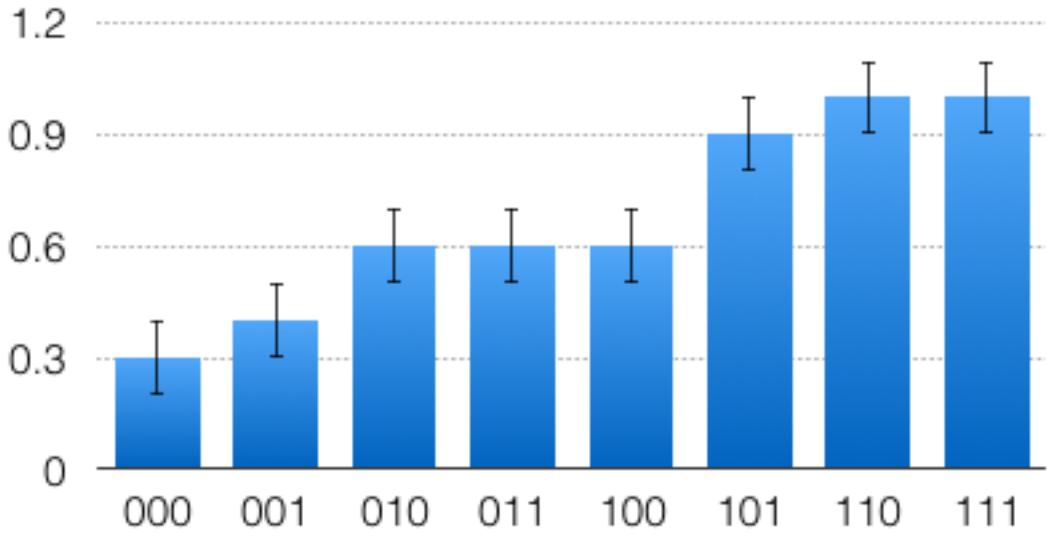
Budget=ε_{global} - ε_1 - ε_2 - ε_3 - ε_4 - ε_5 - ε_6 - ε_7 - ε_8

$X=\{0,1\}^3$ ordered wrt binary encoding.

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$
$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$
$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$
$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$
$q^*_{100}(D) = .6+L(1/n\varepsilon_5)$
$q^*_{101}(D) = .9+L(1/n\varepsilon_6)$
$q^*_{110}(D) = 1+L(1/n\varepsilon_7)$
$q^*_{111}(D) = 1+L(1/n\varepsilon_8)$

$D \in X^{10} =$

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1



Example II

Let's consider the universe domain $\mathcal{X} = \{0,1\}^d$ and let's consider the following predicate for an index $1 \leq j \leq d$

$$q_j(x) = x_j$$

we call an attribute mean function the associated counting query

$$q_j : \mathcal{X}^n \to [0,1]$$

Question: What is the sensitivity?

Example II

Budget=$\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3$

$D \in X^{10} =$

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1
margin	$4+Y_1$	$3+Y_2$	$4+Y_3$

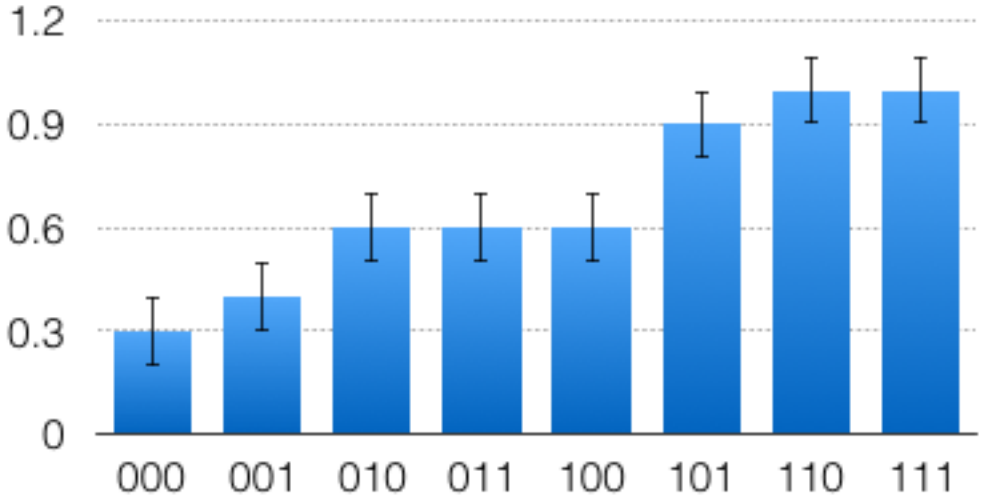
$q^*_1(D) = .4+L(1/n\varepsilon_1)$

$q^*_2(D) = .3+L(1/n\varepsilon_2)$

$q^*_3(D) = .4+L(1/n\varepsilon_3)$

Example II

Budget=ε_{global} - ε_1 - ε_2 - ε_3 - ε_4 - ε_5 - ε_6 - ε_7 - ε_8



Budget=ε_{global} - ε_1 - ε_2 - ε_3

$D \in X^{10} =$

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1
margin	$4+Y_1$	$3+Y_2$	$4+Y_3$

Privacy Budget vs Epsilon

Sometimes is more convenient to think in terms of Privacy Budget: Budget=$\varepsilon_{global} - \sum \varepsilon_{local}$

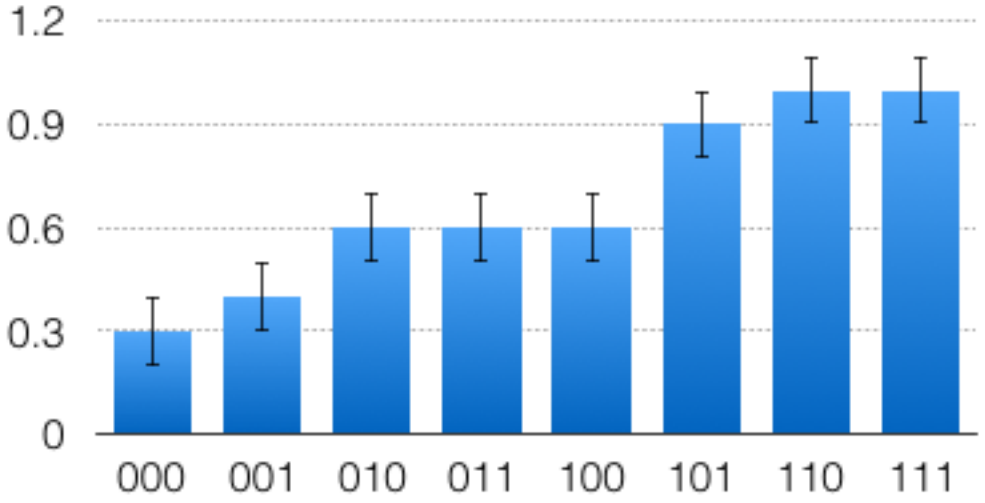
Sometimes is more convenient to think in terms of epsilon: $\varepsilon_{global} = \sum \varepsilon_{local}$

Also making them uniforms is sometimes more informative.

Example II

Budget=ε_{global} - ε_1 - ε_2 - ε_3 - ε_4 - ε_5 - ε_6 - ε_7 - ε_8

$\varepsilon_{global}=\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon=8\varepsilon$



Budget=ε_{global} - ε_1 - ε_2 - ε_3

$\varepsilon_{global}= \varepsilon+\varepsilon+\varepsilon=3\varepsilon$

$D \in X^{10} =$

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1
margin	$4+Y_1$	$3+Y_2$	$4+Y_3$

Composition

Question: How about histograms?

Example III

Let's consider an arbitrary universe domain $\mathcal{X}$ and let's consider the following predicate for $y \in \mathcal{X}$

$$q_y(x) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

we call a point function the associated counting query

$$q_y : \mathcal{X}^n \to [0, 1]$$

Question: What is the sensitivity?

Example III

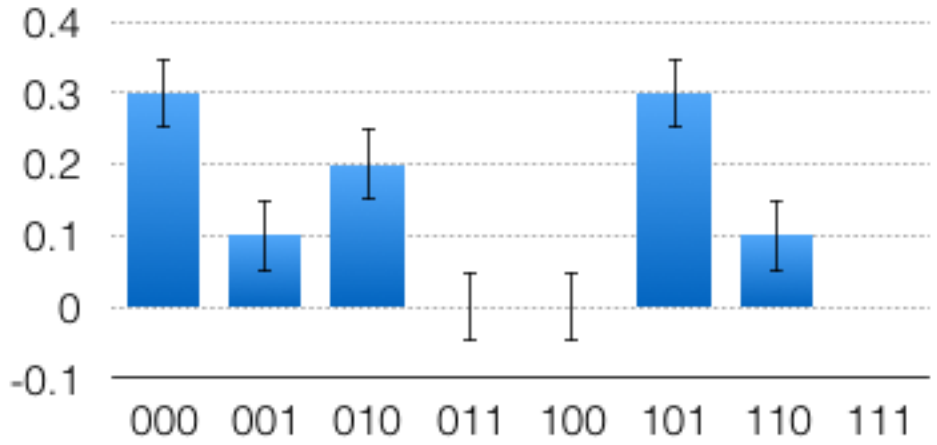
Budget=ε_{global} - ε - ε - ε - ε - ε - ε - ε - ε

Can we do better?

$D \in X^{10}$ =

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1

$q^*_{000}(D) = .3+L(1/n\varepsilon)$
$q^*_{001}(D) = .1+L(1/n\varepsilon)$
$q^*_{010}(D) = .2+L(1/n\varepsilon)$
$q^*_{011}(D) = 0+L(1/n\varepsilon)$
$q^*_{100}(D) = 0+L(1/n\varepsilon)$
$q^*_{101}(D) = .3+L(1/n\varepsilon)$
$q^*_{110}(D) = .1+L(1/n\varepsilon)$
$q^*_{111}(D) = 0+L(1/n\varepsilon)$



Example III

$D \in X^{10} =$

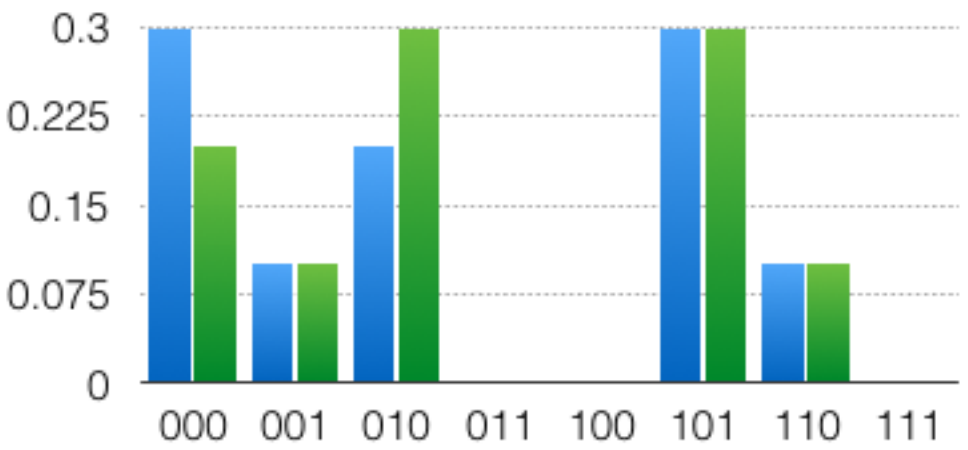
	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1

$D' \in X^{10} =$

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	1	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1

$q_{000}(D) = .3$
$q_{001}(D) = .1$
$q_{010}(D) = .2$
$q_{011}(D) = 0$
$q_{100}(D) = 0$
$q_{101}(D) = .3$
$q_{110}(D) = .1$
$q_{111}(D) = 0$

$q_{000}(D') = .2$
$q_{001}(D') = .1$
$q_{010}(D') = .3$
$q_{011}(D') = 0$
$q_{100}(D') = 0$
$q_{101}(D') = .3$
$q_{110}(D') = .1$
$q_{111}(D') = 0$



Example III

Budget=ε_{global} - 2ε

Can we do better?

$D \in X^{10}$ =

	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1

$q^*_{000}(D) = .3+L(2/n\varepsilon)$
$q^*_{001}(D) = .1+L(2/n\varepsilon)$
$q^*_{010}(D) = .2+L(2/n\varepsilon)$
$q^*_{011}(D) = 0+L(2/n\varepsilon)$
$q^*_{100}(D) = 0+L(2/n\varepsilon)$
$q^*_{101}(D) = .3+L(2/n\varepsilon)$
$q^*_{110}(D) = .1+L(2/n\varepsilon)$
$q^*_{111}(D) = 0+L(2/n\varepsilon)$