Differential Privacy

Basic properties.

Marco Gaboardi
Boston University

The opinions expressed in this course are mine and they do not not reflect those of the National Science Foundation or the US. Census Bureau.

(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 $S \subseteq R$:

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

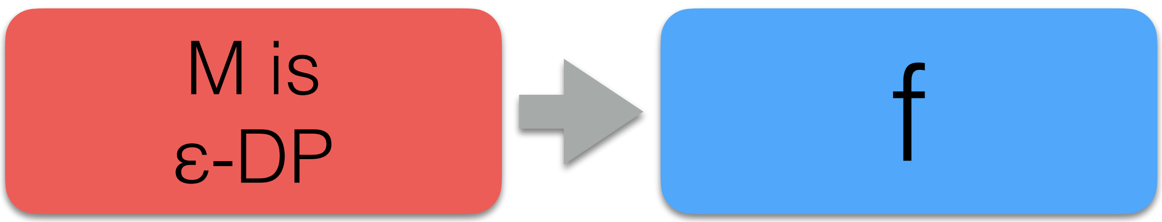
Some important properties

- Resilience to post-processing
- Group privacy
- Composition

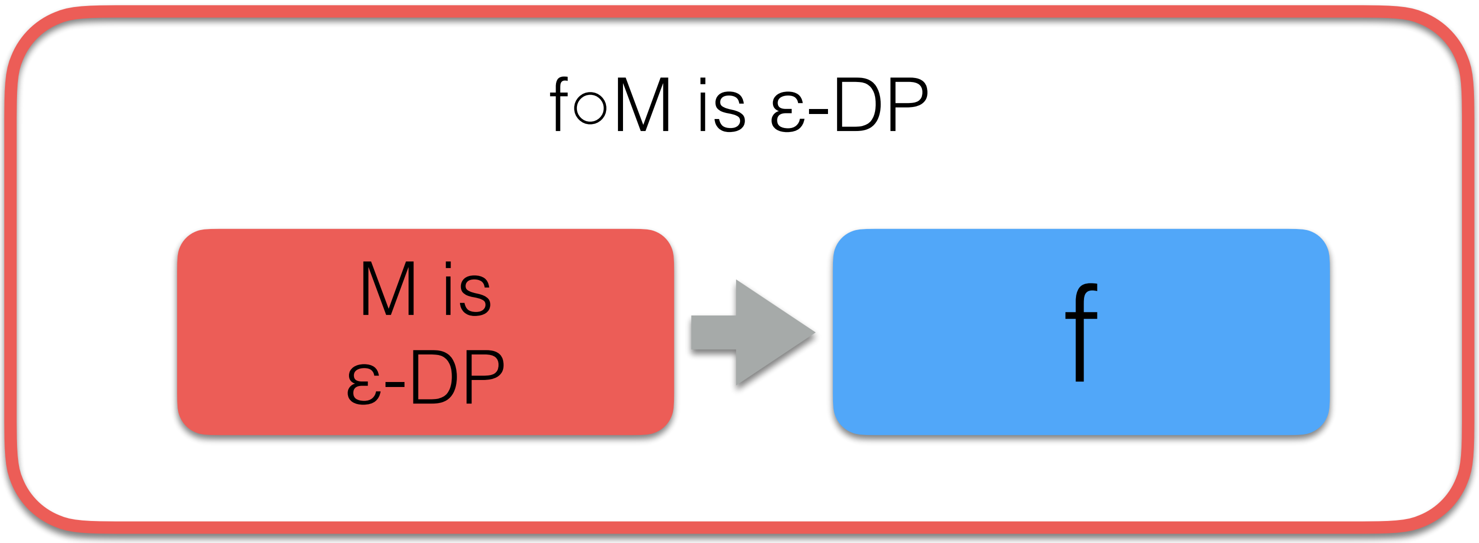
Resilience to Post-processing

M is
ε-DP

Resilience to Post-processing



Resilience to Post-processing



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.

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.

Proof. Fix any pair of neighboring databases $D \sim_1 D'$, and fix any event $S \subseteq R'$. Let $T = \{r \in R : f(r) \in S\}$. We have

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.

Proof. Fix any pair of neighboring databases $D \sim_1 D'$, and fix any event $S \subseteq R'$. Let $T = \{r \in R : f(r) \in S\}$. We have

$$\Pr[f(\mathcal{M}(D)) \in S] \quad = \quad \Pr[\mathcal{M}(D) \in T]$$

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.

Proof. Fix any pair of neighboring databases $D \sim_1 D'$, and fix any event $S \subseteq R'$. Let $T = \{r \in R : f(r) \in S\}$. We have

$$\begin{aligned}\Pr[f(\mathcal{M}(D)) \in S] &= \Pr[\mathcal{M}(D) \in T] \\ &\leq \exp(\epsilon) Pr[\mathcal{M}(D') \in T]\end{aligned}$$

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.

Proof. Fix any pair of neighboring databases $D \sim_1 D'$, and fix any event $S \subseteq R'$. Let $T = \{r \in R : f(r) \in S\}$. We have

$$\begin{aligned} \Pr[f(\mathcal{M}(D)) \in S] &= \Pr[\mathcal{M}(D) \in T] \\ &\leq \exp(\epsilon) Pr[\mathcal{M}(D') \in T] \\ &= \exp(\epsilon) Pr[f(\mathcal{M}(D')) \in S] \end{aligned}$$

Resilience to Post-processing

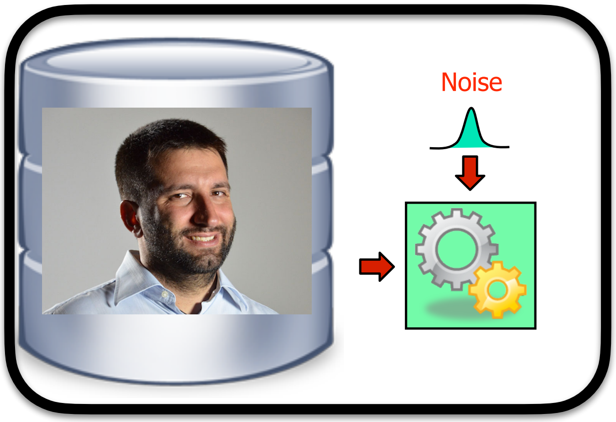
Question: Why is resilience to post-processing important?

Resilience to Post-processing

Question: Why is resilience to post-processing important?

Answer: Because it is what allows us to publicly release the result of a differentially private analysis!

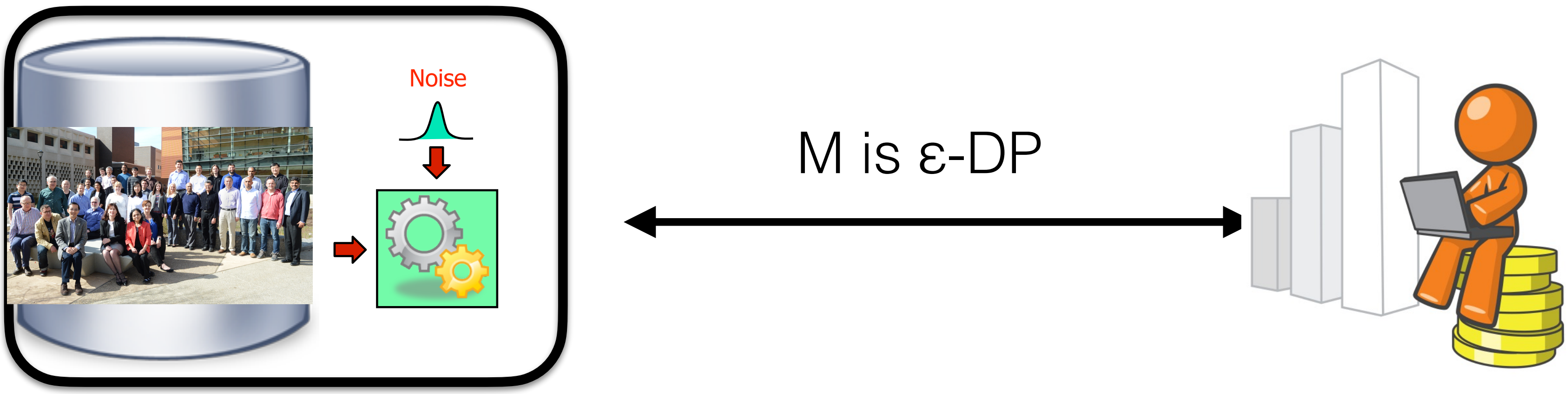
Group Privacy



M is ε-DP

$$\Pr[\mathcal{M}(D) = r] \leq e^{\epsilon} \Pr[\mathcal{M}(D') = r]$$

Group Privacy



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

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

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

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

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

Proof. Fix any pair of databases D, D' with $D \Delta D' \leq k$. Then, we have databases $D_0, D_1, \ldots, D_k$ such that $D_0 = D$, $D_k = D'$ and $D_i \Delta D_{i+1} \leq 1$. Fix also any event $S \subseteq R'$. Then, we have have

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

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

Proof. Fix any pair of databases D, D' with $D \Delta D' \leq k$. Then, we have databases $D_0, D_1, \ldots, D_k$ such that $D_0 = D$, $D_k = D'$ and $D_i \Delta D_{i+1} \leq 1$. Fix also any event $S \subseteq R'$. Then, we have have

$$\begin{aligned} \Pr[\mathcal{M}(D) \in S] &= \Pr[\mathcal{M}(D_0) \in S] \\ &\leq \exp(\epsilon) \Pr[\mathcal{M}(D_1) \in S] \end{aligned}$$

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

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

Proof. Fix any pair of databases D, D' with $D \Delta D' \leq k$. Then, we have databases $D_0, D_1, \ldots, D_k$ such that $D_0 = D$, $D_k = D'$ and $D_i \Delta D_{i+1} \leq 1$. Fix also any event $S \subseteq R'$. Then, we have have

$$\begin{aligned} \Pr[\mathcal{M}(D) \in S] &= \Pr[\mathcal{M}(D_0) \in S] \\ &\leq \exp(\epsilon) \Pr[\mathcal{M}(D_1) \in S] \\ &\leq \exp(\epsilon)(\exp(\epsilon) \Pr[\mathcal{M}(D_2) \in S]) = \exp(2\epsilon) \Pr[\mathcal{M}(D_2) \in S] \end{aligned}$$

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

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

Proof. Fix any pair of databases D, D' with $D\Delta D' \leq k$. Then, we have databases $D_0, D_1, \ldots, D_k$ such that $D_0 = D$, $D_k = D'$ and $D_i \Delta D_{i+1} \leq 1$. Fix also any event $S \subseteq R'$. Then, we have have

$$\begin{aligned}
\Pr[\mathcal{M}(D) \in S] &= \Pr[\mathcal{M}(D_0) \in S] \\
&\leq \exp(\epsilon) \Pr[\mathcal{M}(D_1) \in S] \\
&\leq \exp(\epsilon)(\exp(\epsilon) \Pr[\mathcal{M}(D_2) \in S]) = \exp(2\epsilon) \Pr[\mathcal{M}(D_2) \in S] \\
&\leq \cdots \\
&\leq \exp(k\epsilon) \Pr[\mathcal{M}(D_k) \in S] = \exp(k\epsilon) \Pr[\mathcal{M}(D') \in S]
\end{aligned}$$

Group Privacy

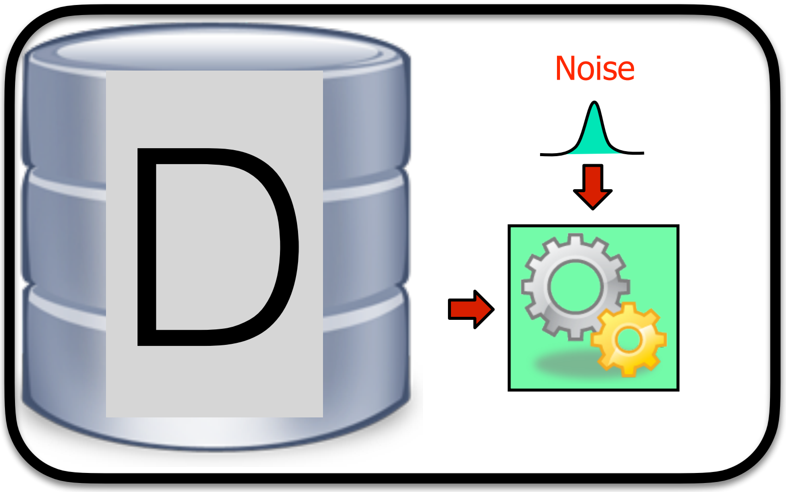
Question: Why is group privacy important?

Group Privacy

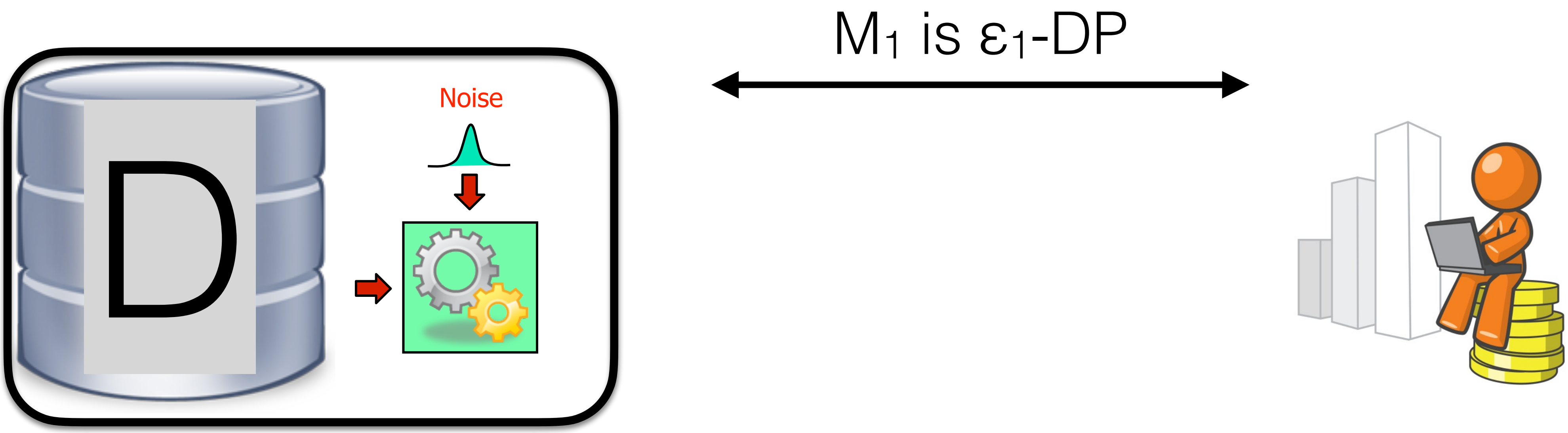
Question: Why is group privacy important?

Answer: Because it allows to reason about privacy at different level of granularities!

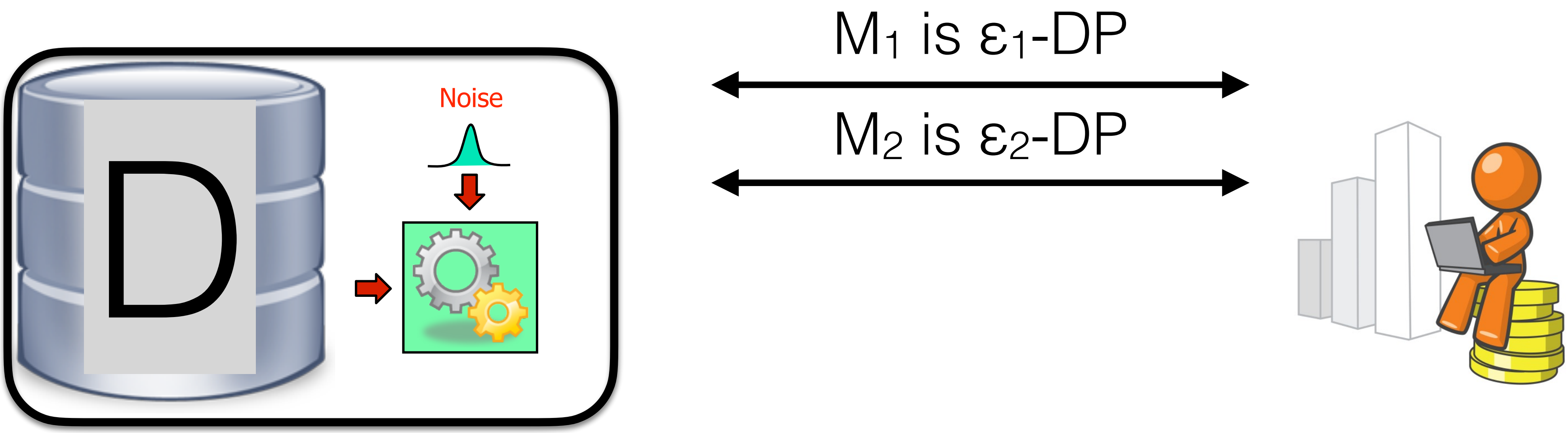
Composition



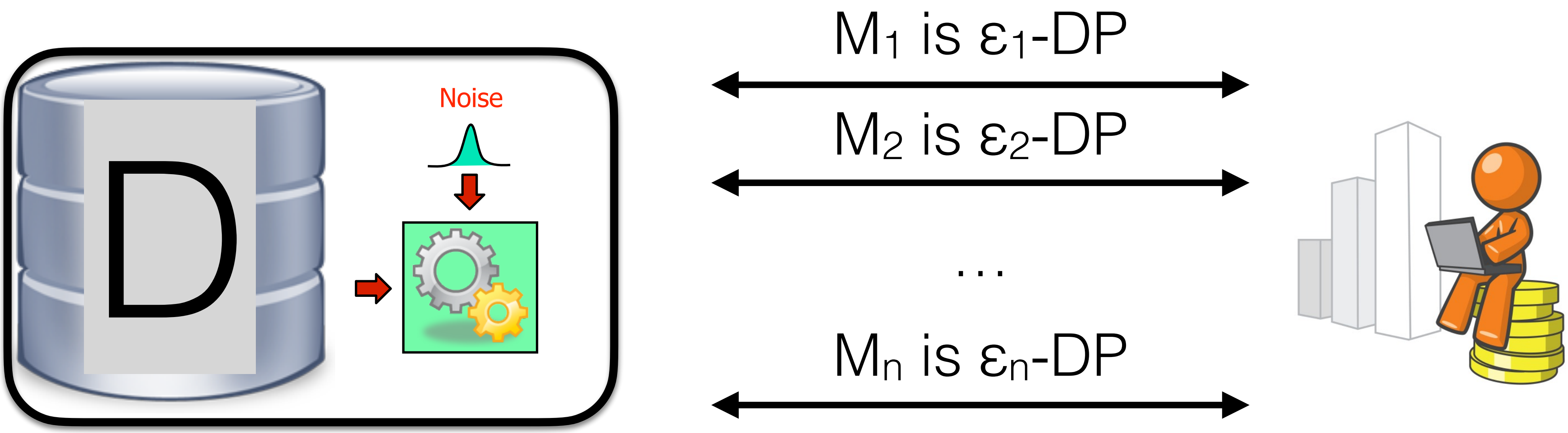
Composition



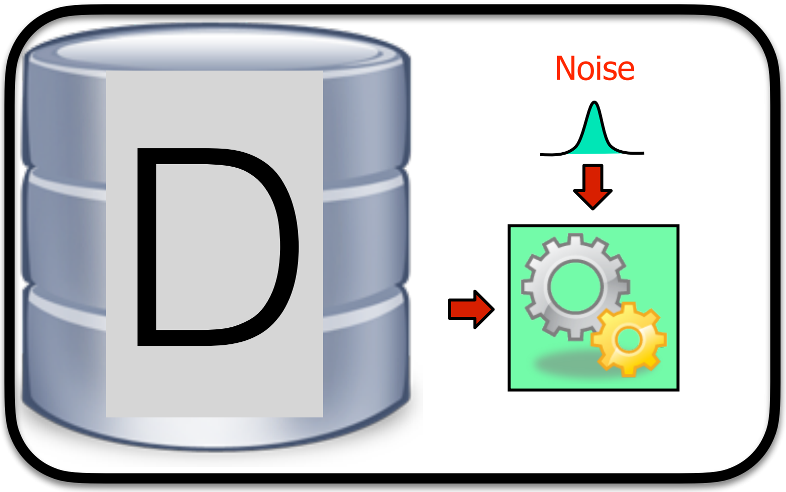
Composition



Composition



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.

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:

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

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:

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

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

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:

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

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

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:

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

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

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

Composition

Question: Why composition is important?

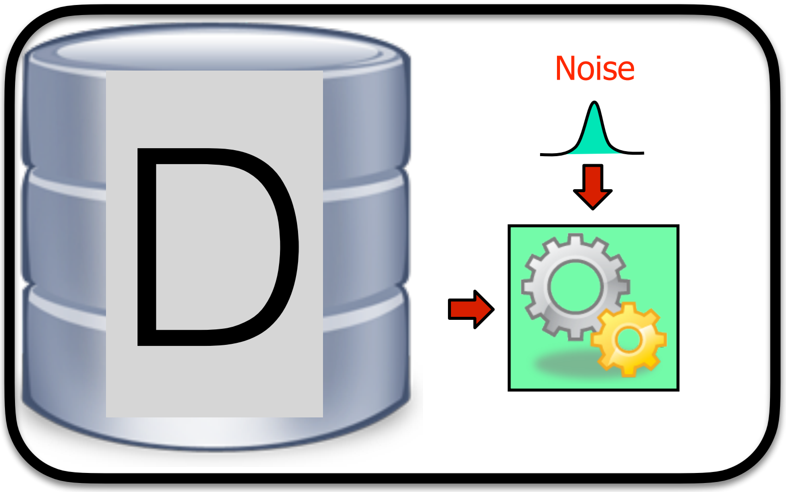
Composition

Question: Why composition is important?

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

Composition

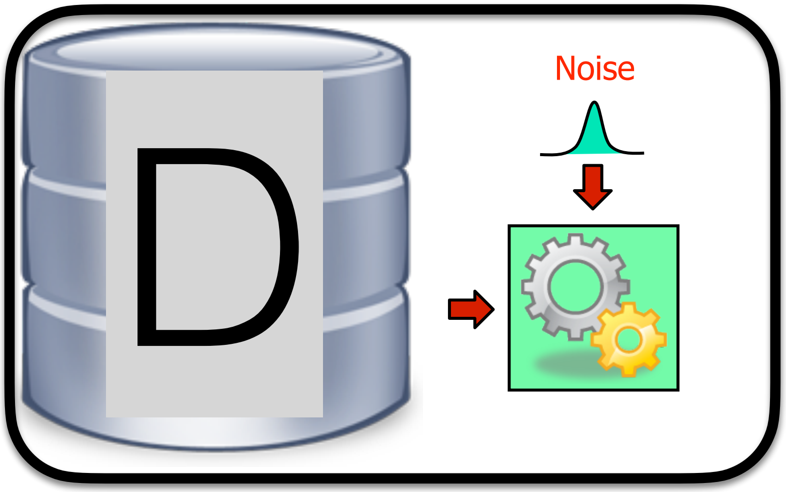
Budget=ε_{global}



Composition

Budget=ε_{global}

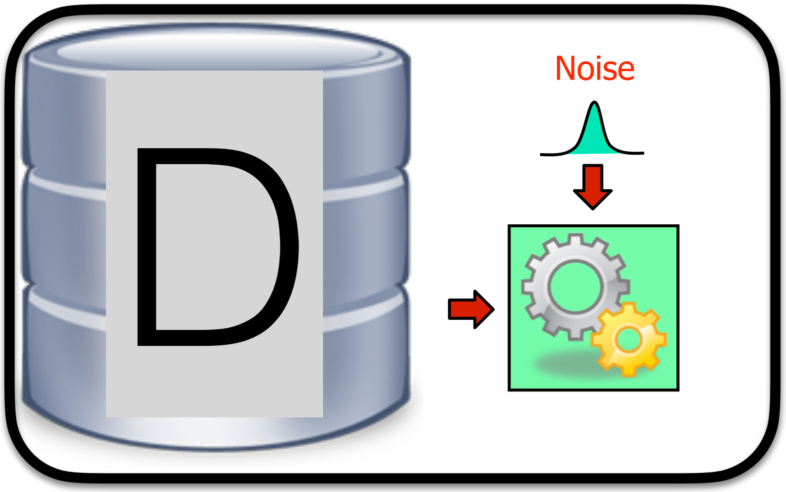
M_1 is ε_1-DP

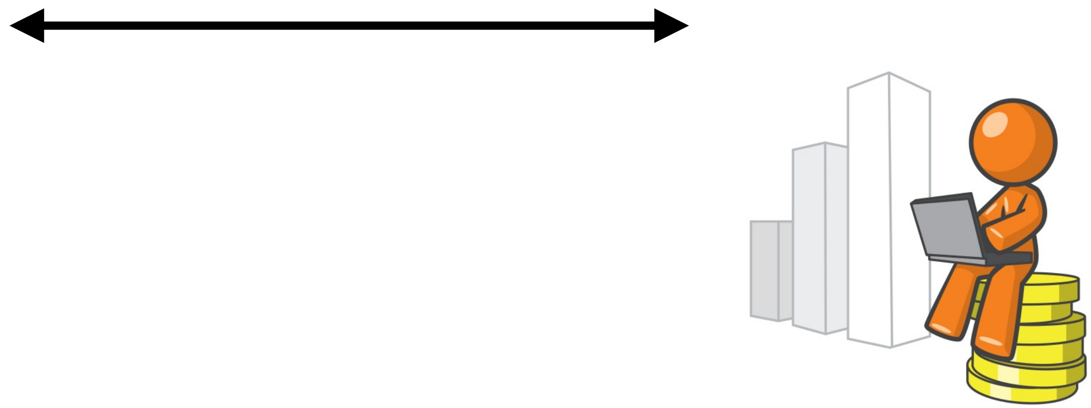


Composition

Budget=ε_{global} - ε_1

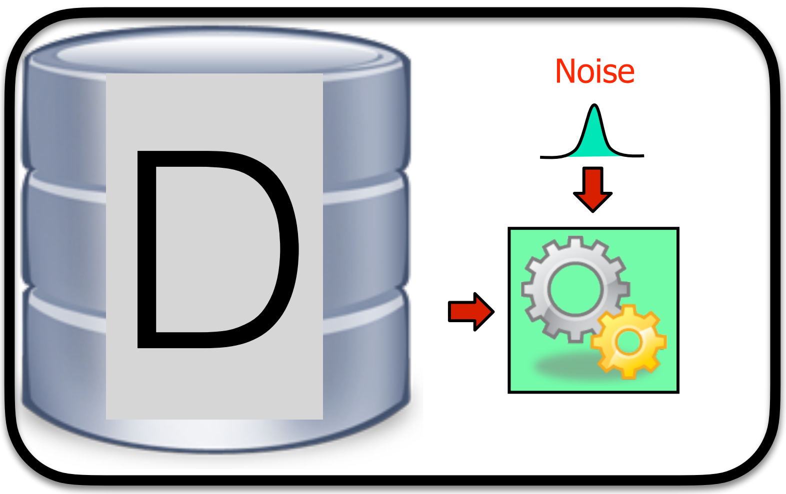
M_1 is ε_1-DP





Composition

Budget=ε_{global} - ε_1

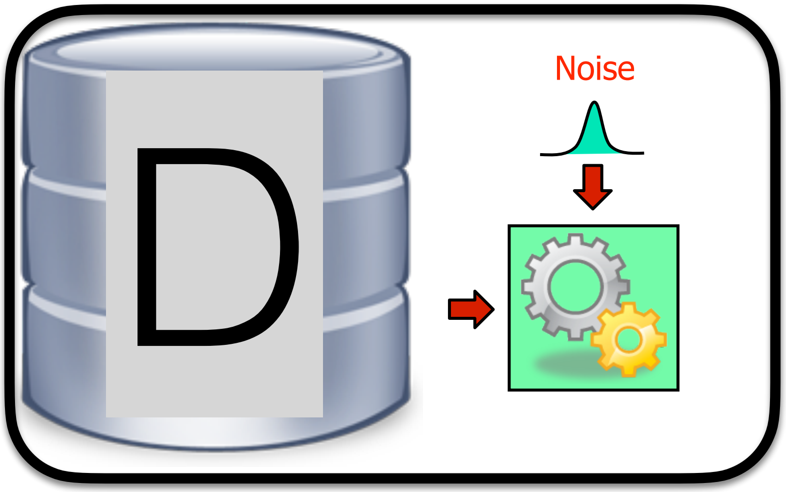


M_1 is ε_1-DP

M_2 is ε_2-DP

Composition

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

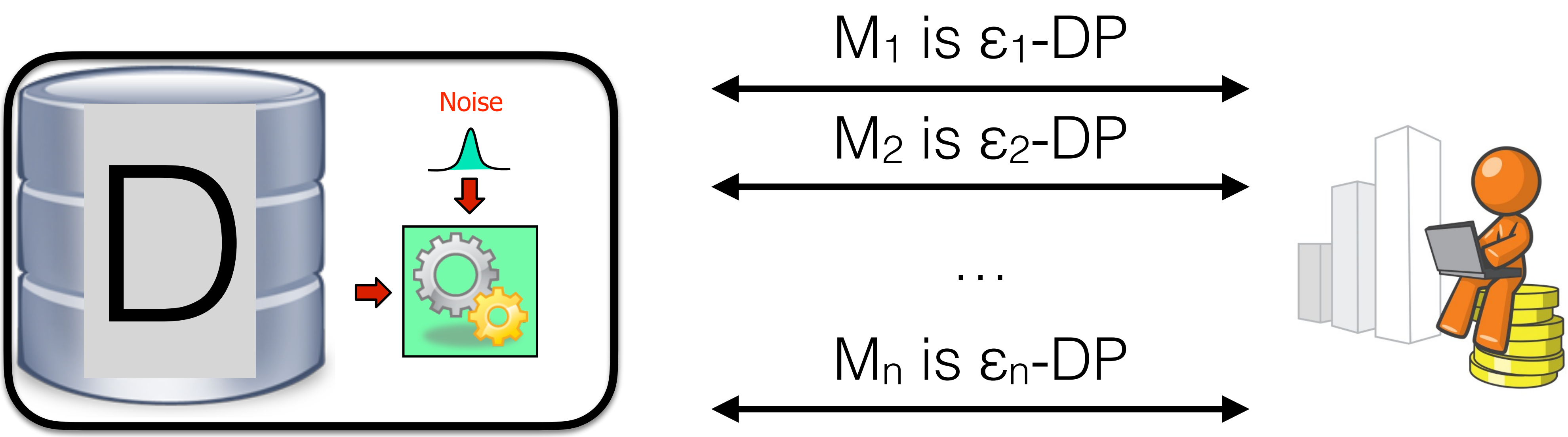


M_1 is ε_1-DP

M_2 is ε_2-DP

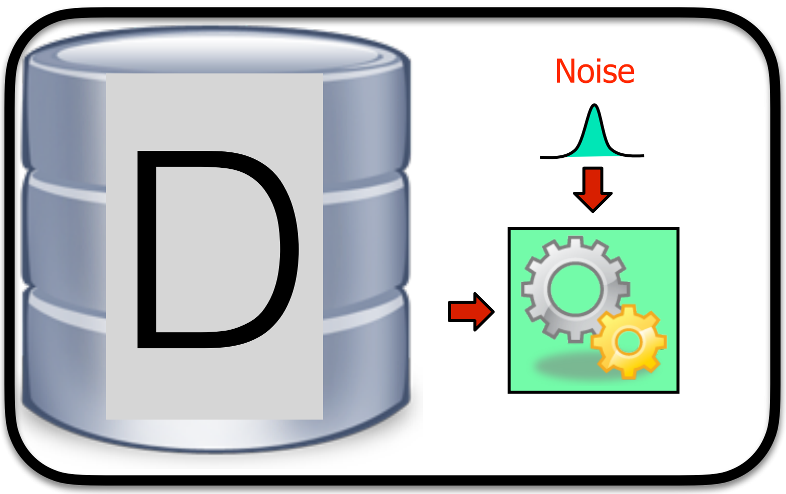
Composition

Budget=ε_{global} - ε_1 - ε_2 ...



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

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

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 1

Budget=ε_{global}

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

$D \in X^{10} =$

	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 I

$\text{Budget} = \varepsilon_{global} - \varepsilon_1$

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

$q^*_{000}(D) = .3 + L(1/n\varepsilon_1)$

$D \in X^{10} =$

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

Example 1

Budget=ε_{global} - ε_1 - ε_2

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

$q^*_{000}(D) = .3 + L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4 + L(1/n\varepsilon_2)$

$D \in X^{10} =$

	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 I

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

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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

$D \in X^{10} =$

	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 I

Budget=$\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4$

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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$

$D \in X^{10} =$

	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 1

Budget=$\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 - \varepsilon_5$

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

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$

$q^*_{100}(D) = .6+L(1/n\varepsilon_5)$

Example I

Budget=$\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 - \varepsilon_5 - \varepsilon_6$

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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$

$q^*_{100}(D) = .6+L(1/n\varepsilon_5)$

$q^*_{101}(D) = .9+L(1/n\varepsilon_6)$

$D \in X^{10} =$

	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 I

Budget=$\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 - \varepsilon_5 - \varepsilon_6 - \varepsilon_7$

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

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$

$q^*_{100}(D) = .6+L(1/n\varepsilon_5)$

$q^*_{101}(D) = .9+L(1/n\varepsilon_6)$

$q^*_{110}(D) = 1+L(1/n\varepsilon_7)$

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.

$D \in X^{10}$ =

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$

$q^*_{100}(D) = .6+L(1/n\varepsilon_5)$

$q^*_{101}(D) = .9+L(1/n\varepsilon_6)$

$q^*_{110}(D) = 1+L(1/n\varepsilon_7)$

$q^*_{111}(D) = 1+L(1/n\varepsilon_8)$

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.

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon_1)$

$q^*_{001}(D) = .4+L(1/n\varepsilon_2)$

$q^*_{010}(D) = .6+L(1/n\varepsilon_3)$

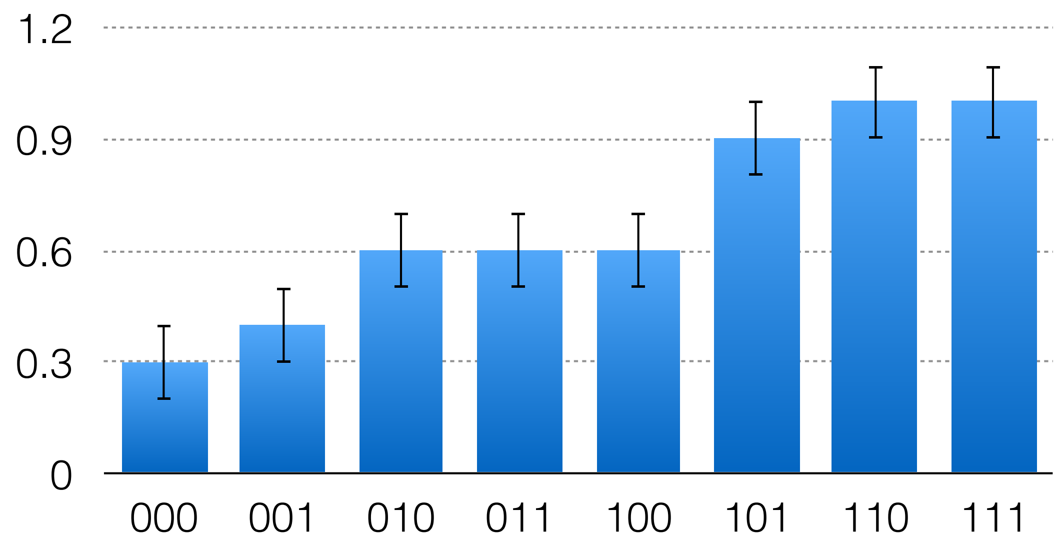
$q^*_{011}(D) = .6+L(1/n\varepsilon_4)$

$q^*_{100}(D) = .6+L(1/n\varepsilon_5)$

$q^*_{101}(D) = .9+L(1/n\varepsilon_6)$

$q^*_{110}(D) = 1+L(1/n\varepsilon_7)$

$q^*_{111}(D) = 1+L(1/n\varepsilon_8)$



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

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

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

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

Question: What is the sensitivity?

Example II

Budget=ε_{global}

$D \in X^{10} =$

	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

Budget=ε_{global} - ε_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	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1

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

Example II

Budget=ε_{global} - ε_1 - ε_2

$D \in X^{10} =$

	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

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

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

Example II

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

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

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

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

Example II

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

$D \in X^{10} =$

	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	$4+Y_1$	$3+Y_2$	$4+Y_3$

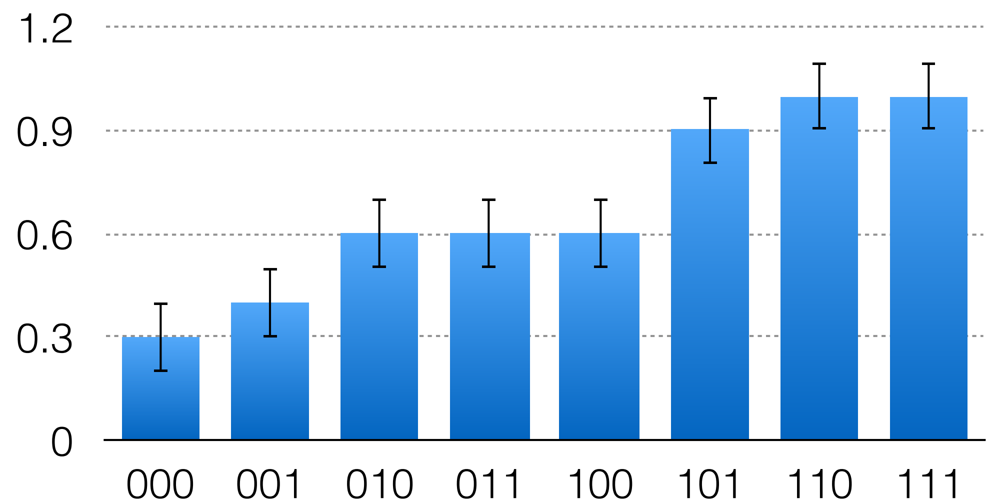
$q^*_1(D) = .4+L(1/n\varepsilon_1)$

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

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

Example II

Budget= $\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 - \varepsilon_5 - \varepsilon_6 - \varepsilon_7 - \varepsilon_8$



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	$4+Y_1$	$3+Y_2$	$4+Y_3$

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: $\varepsilon_{global} = \sum \varepsilon_{local}$

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: $\varepsilon_{global} = \sum \varepsilon_{local}$

Making them uniforms is sometimes more informative.

Privacy Budget vs Epsilon

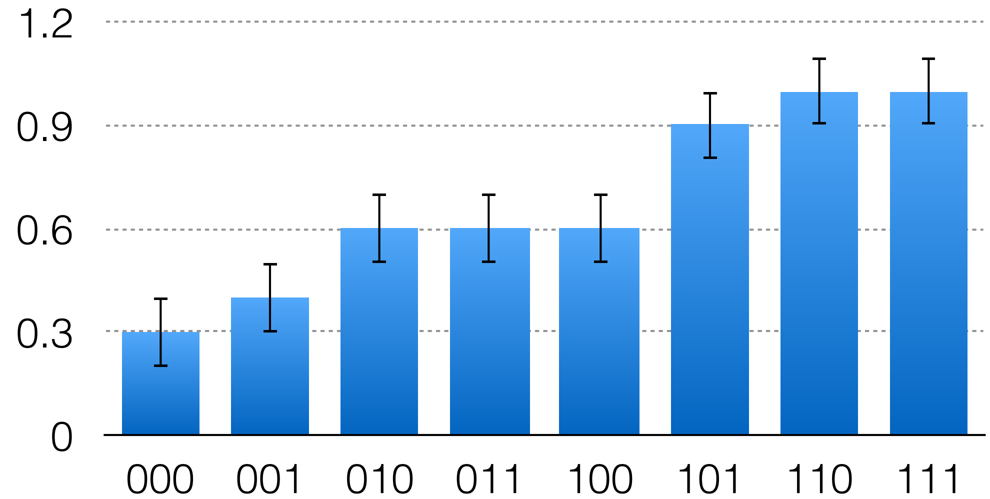
Sometimes is more convenient to think in terms of Privacy Budget: Budget=$\varepsilon_{global} - \sum \varepsilon_{local}$

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

Making them uniforms is sometimes more informative.

Note: There are situations where the two are not equivalent.

Example II



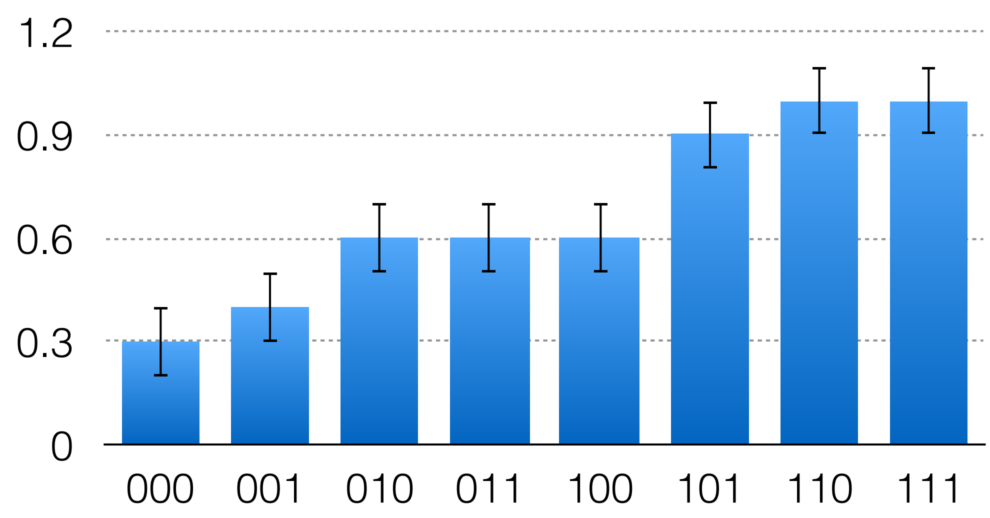
$\text{Budget} = \varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 - \varepsilon_5 - \varepsilon_6 - \varepsilon_7 - \varepsilon_8$

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

$D \in X^{10} =$

	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	$4+Y_1$	$3+Y_2$	$4+Y_3$

Example II



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

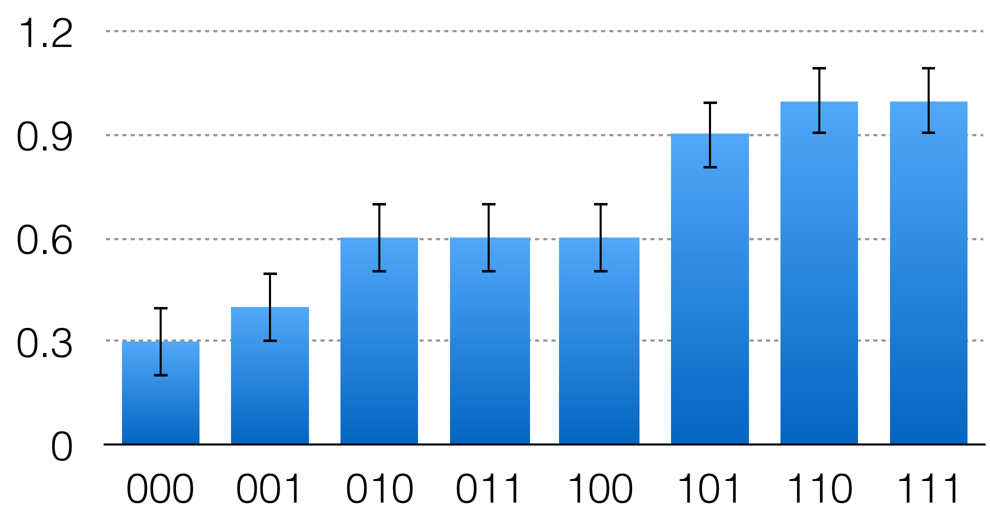
$\varepsilon_{global} = \varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon+\varepsilon = 8\varepsilon$

Budget= ε_{global} - ε_1 - ε_2 - ε_3

$D \in X^{10} =$

	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	$4+Y_1$	$3+Y_2$	$4+Y_3$

Example II



Budget= $\varepsilon_{global} - \varepsilon_1 - \varepsilon_2 - \varepsilon_3 - \varepsilon_4 - \varepsilon_5 - \varepsilon_6 - \varepsilon_7 - \varepsilon_8$

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

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

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

$D \in X^{10} =$

	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

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

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}

$D \in X^{10} =$

	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 III

Budget=ε_{global} - ε

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

Example III

Budget=ε_{global} - ε - ε

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

Example III

$\text{Budget} = \varepsilon_{global} - \varepsilon \quad - \varepsilon \quad - \varepsilon$

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3 + L(1/n\varepsilon)$

$q^*_{001}(D) = .1 + L(1/n\varepsilon)$

$q^*_{010}(D) = .2 + L(1/n\varepsilon)$

Example III

Budget=ε_{global} - ε - ε - ε - ε

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

$q^*_{010}(D) = .2+L(1/n\varepsilon)$

$q^*_{011}(D) = 0+L(1/n\varepsilon)$

Example III

Budget=ε_{global} - ε - ε - ε - ε - ε

$D \in X^{10}$ =

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

$q^*_{010}(D) = .2+L(1/n\varepsilon)$

$q^*_{011}(D) = 0+L(1/n\varepsilon)$

$q^*_{100}(D) = 0+L(1/n\varepsilon)$

Example III

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

$D \in X^{10}$ =

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

$q^*_{010}(D) = .2+L(1/n\varepsilon)$

$q^*_{011}(D) = 0+L(1/n\varepsilon)$

$q^*_{100}(D) = 0+L(1/n\varepsilon)$

$q^*_{101}(D) = .3+L(1/n\varepsilon)$

Example III

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

$D \in X^{10} =$

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

$q^*_{010}(D) = .2+L(1/n\varepsilon)$

$q^*_{011}(D) = 0+L(1/n\varepsilon)$

$q^*_{100}(D) = 0+L(1/n\varepsilon)$

$q^*_{101}(D) = .3+L(1/n\varepsilon)$

$q^*_{110}(D) = .1+L(1/n\varepsilon)$

Example III

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

$D \in X^{10} =$

	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

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

$D \in X^{10}$ =

	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

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

$q^*_{010}(D) = .2+L(1/n\varepsilon)$

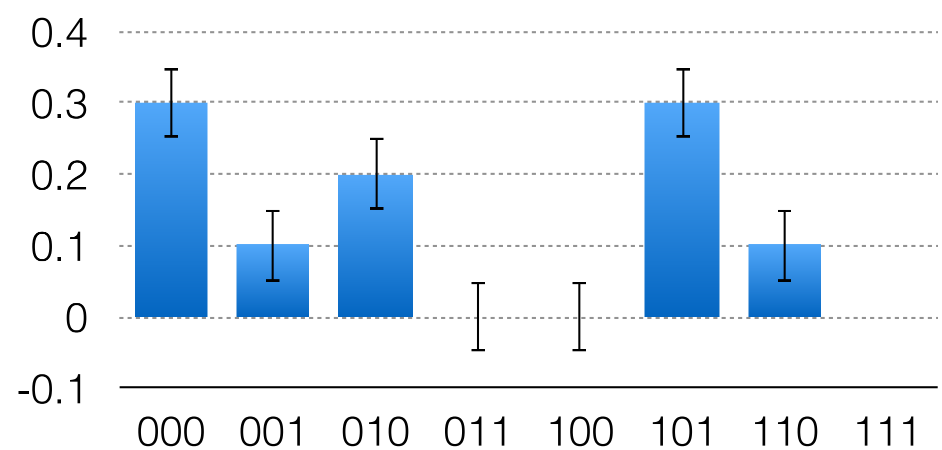
$q^*_{011}(D) = 0+L(1/n\varepsilon)$

$q^*_{100}(D) = 0+L(1/n\varepsilon)$

$q^*_{101}(D) = .3+L(1/n\varepsilon)$

$q^*_{110}(D) = .1+L(1/n\varepsilon)$

$q^*_{111}(D) = 0+L(1/n\varepsilon)$



Example III

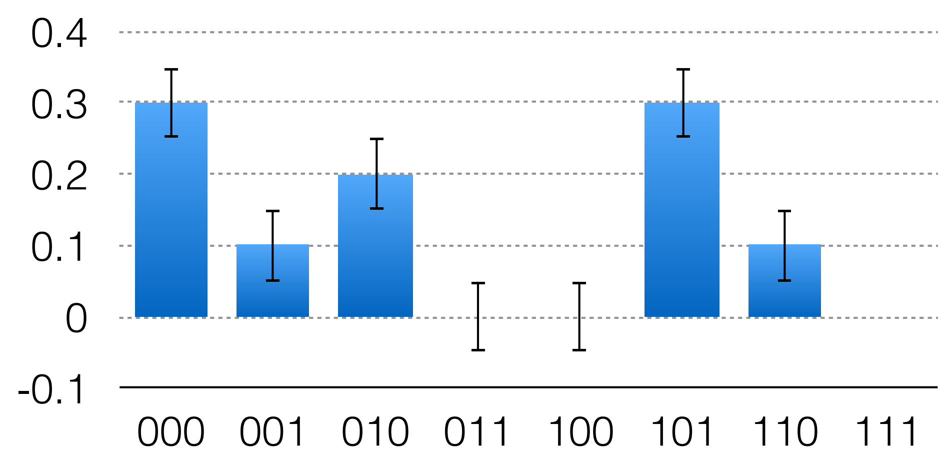
Budget=ε_{global} - ε - ε - ε - ε
- ε - ε - ε - ε

Can we do better?

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

$D \in X^{10} =$

	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 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

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

$q_{000}(D) = .3$

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

$q_{000}(D) = .3$
$q_{001}(D) = .1$

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

$q_{000}(D) = .3$

$q_{001}(D) = .1$

$q_{010}(D) = .2$

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

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

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

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

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

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

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

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

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

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

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

$q_{000}(D) = .3$

$q_{001}(D) = .1$

$q_{010}(D) = .2$

$q_{011}(D) = 0$

$q_{100}(D) = 0$

$q_{101}(D) = .3$

$q_{110}(D) = .1$

$q_{111}(D) = 0$

$q_{000}(D') = .2$

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

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

$q_{000}(D') = .2$
$q_{001}(D') = .1$

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

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

$q_{000}(D') = .2$
$q_{001}(D') = .1$
$q_{010}(D') = .3$

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

$D \in X^{10} =$

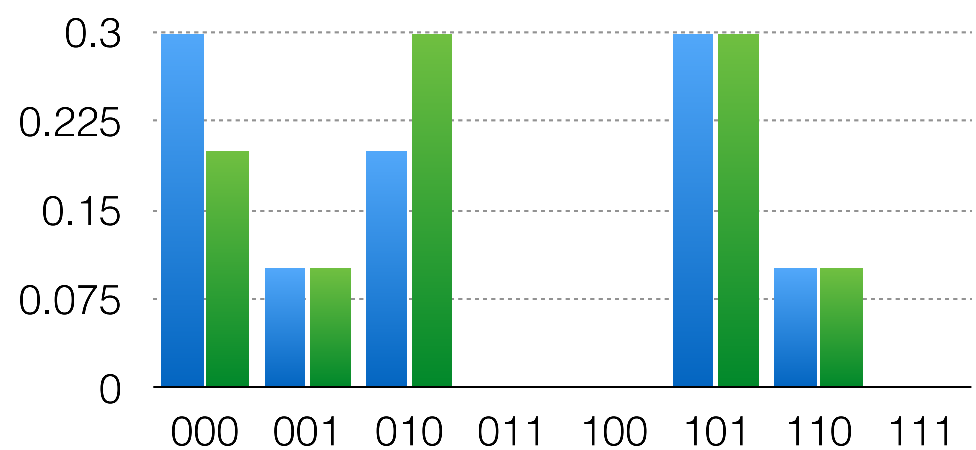
	D1	D2	D3
I1	0	0	0
I2	1	0	1
I3	0	1	0
I4	1	0	1
I5	0	0	0
I6	0	0	1
I7	1	1	0
I8	0	0	0
I9	0	1	0
I10	1	0	1

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

	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?

$q^*_{000}(D) = .3+L(1/n\varepsilon)$

$q^*_{001}(D) = .1+L(1/n\varepsilon)$

$q^*_{010}(D) = .2+L(1/n\varepsilon)$

$q^*_{011}(D) = 0+L(1/n\varepsilon)$

$q^*_{100}(D) = 0+L(1/n\varepsilon)$

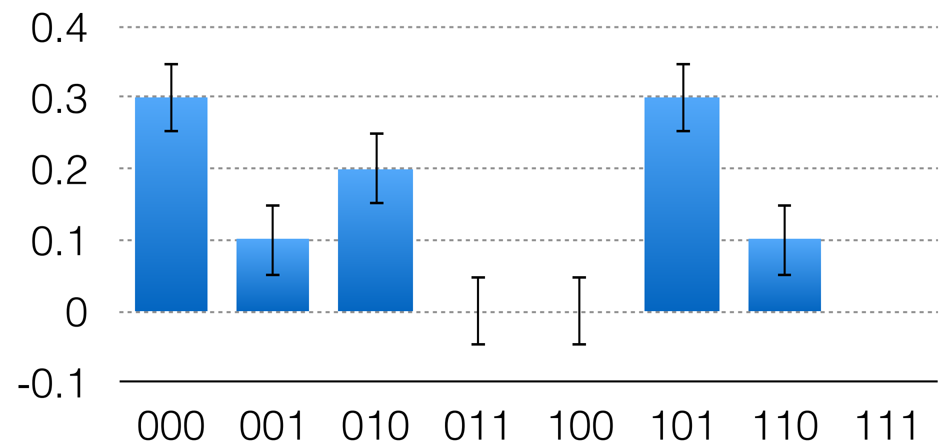
$q^*_{101}(D) = .3+L(1/n\varepsilon)$

$q^*_{110}(D) = .1+L(1/n\varepsilon)$

$q^*_{111}(D) = 0+L(1/n\varepsilon)$

$D \in X^{10} =$

	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



Releasing partial sums

```
DummySum(d : {0,1} list) : real list
  i:= 0;
  s:= 0;
  r:= [];
  while (i<size d)
     s:= s + d[i]
     z:=$ s + Lap(1/eps)
     r:= r ++ [z];
     i:= i+1;
  return r
```

What is the global epsilon here?

Releasing partial sums

```
DummySum(d : {0,1} list) : real list
  i:=0;
  s:=0;
  r:=[];
  while (i<size d)
     z:=$ d[I] + Lap(eps)
     s:= s + z
     r:= r ++ [s];
     i:= i+1;
  return r
```

What is the global epsilon here?

Parallel Composition

Let M_1:DB →R be a (ε_1,δ_1)-differentially private program and M_2:DB →R be a (ε_2,δ_2)-differentially private program. Suppose that we partition D in a data-independent way into two datasets D_1 and D_2. Then, the composition $M_{1,2}$:DB→R defined as

$$MP_{1,2}(D)=(M_1(D_1),M_2(D_2))$$

is $(\max(\varepsilon_1,\varepsilon_2),\max(\delta_1,\delta_2))$-differentially private.

Composition

Question: how much perturbation do we have if we want to answer n queries under ε-DP?

Composition

Question: how much perturbation do we have if we want to answer n counting queries under ε_{global}-DP?

We can split the privacy budget uniformly:

$$\epsilon = \frac{\epsilon_{\text{global}}}{n}$$

Composition

Question: how much perturbation do we have if we want to answer n counting queries under ε_{global}-DP?

We can split the privacy budget uniformly:

$$\epsilon = \frac{\epsilon_{\text{global}}}{n}$$

Laplace accuracy: with high probability we have:

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

Composition

Question: how much perturbation do we have if we want to answer n counting queries under $\varepsilon_{\text{global}}$-DP?

By putting them together (hiding some details) we have as a max error

$$O\Big(\frac{n}{\epsilon_{\text{global}} n}\Big) = O\Big(\frac{1}{\epsilon_{\text{global}}}\Big)$$

Composition

Question: how much perturbation do we have if we want to answer n counting queries under ε_{global}-DP?

By putting them together (hiding some details) we have as a max error

$$O\Big(\frac{n}{\epsilon_{\text{global}} n}\Big) = O\Big(\frac{1}{\epsilon_{\text{global}}}\Big)$$

Notice that if we don't renormalize this is of the order of

$$O\Big(\frac{n}{\epsilon_{\text{global}}}\Big)$$

bigger than the sample error.

Composition

Question: how many counting queries can we answer with small error under ε_{global}-DP?

Let's now target an error similar to sample error. How many queries we can answer?

If we want a non-normalized error of:

$$O\left(\frac{\sqrt{n}}{\epsilon_{\text{global}}}\right)$$

we can answer at most $\sqrt{n}$ queries.

Composition

Question: Can we do better?

Composition

Question: how much perturbation do we have if we want to answer n queries under (ε,δ)-DP?

Composition

Question: how much perturbation do we have if we want to answer n queries under (ε,δ)-DP?

We have (by hiding many details) as a max error

$$O\Big(\frac{1}{\epsilon_{\text{global}}\sqrt{n}}\Big)$$

[DworkRothblumVadhan10, SteinkeUllman16]

Composition

Question: how much perturbation do we have if we want to answer n queries under (ε,δ)-DP?

We have (by hiding many details) as a max error

$$O\left(\frac{1}{\epsilon_{\text{global}}\sqrt{n}}\right)$$

If we don't renormalize this is of the order of

$$O\left(\frac{\sqrt{n}}{\epsilon_{\text{global}}}\right)$$

comparable to the sample error.

[DworkRothblumVadhan10, SteinkeUllman16]

Summary

- Resilience to post-processing
- Group privacy
- Composition