Extremal Mechanisms for Local Differential Privacy

Peter Kairouz

Department of Electrical & Computer Engineering
University of Illinois at Urbana-Champaign

Joint work with Sewoong Oh (UIUC) and Pramod Viswanath (UIUC)

Private Communication vs. Secure Communication

The fundamental limits of digital communication are well understood

Private Communication vs. Secure Communication

Secure communication is a fairly mature technology

Private Communication vs. Secure Communication

The **fundamental limits** of **privacy** have not been explored yet

Private Communication vs. Secure Communication

We study the **fundamental** trade-off between **privacy** and **utility**

Does Privacy Matter? [Greenwald 2014]

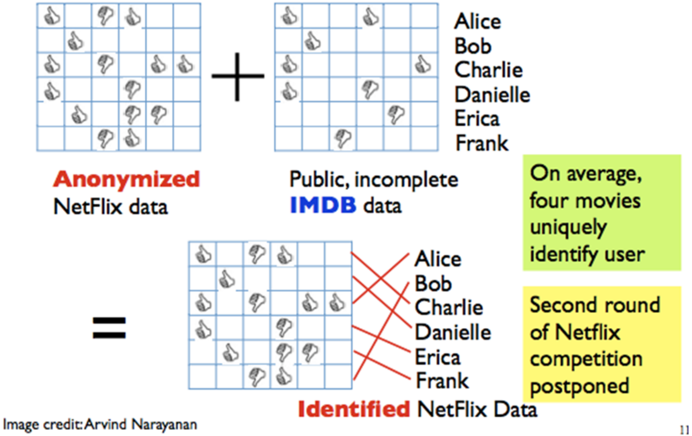
"If you're doing something that you don't want other people to know, maybe you shouldn't be doing it in first place"

"Privacy is no longer a social norm!"

Recent Privacy Leaks

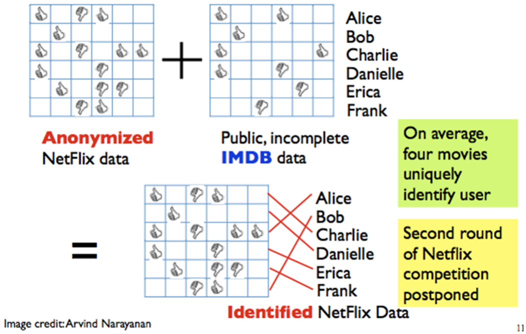
From **anonymous faces** to **social security numbers**

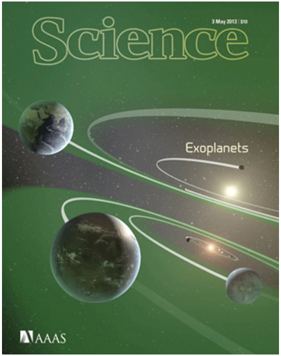
Recent Privacy Leaks



Deanonymizing Netflix data

Recent Privacy Leaks





Deanonymizing Netflix data, **identifying** personal genomes, etc.

Privacy is a **fundamental** human **right!**

The Ultimate Protection

"The future of privacy is **lying**"

randomizing = **systematic lying**

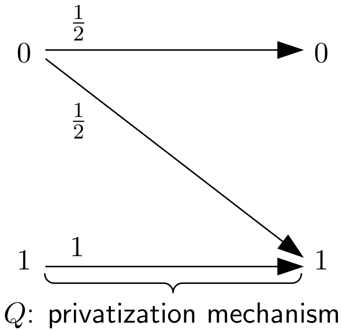
Privacy via Plausible Deniability [Warner 1965]

Have you ever used illegal drugs?

say **yes**

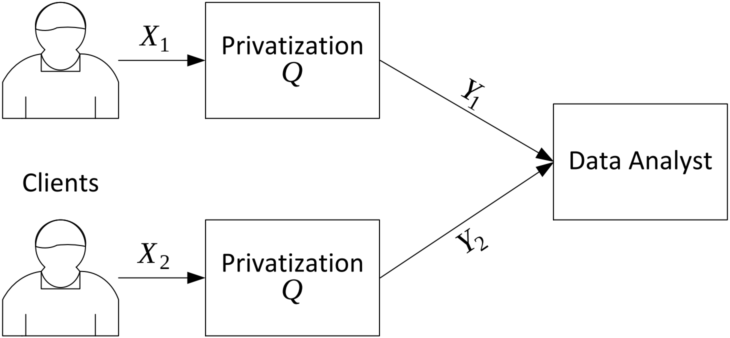
answer **truthfully**

Privacy via Plausible Deniability [Warner 1965]



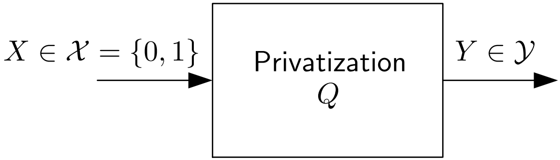
- instead of $X = x$, share $Y = y$ w.p. $Q\left(y|x\right)$
- $Q : |\mathcal{X}| \times |\mathcal{Y}|$ stochastic mapping

The Local Privacy Model [Duchi, et. al., 2012]



- clients **receive a service** if they share their data
- clients **do not trust** data analysts

Inference of Information

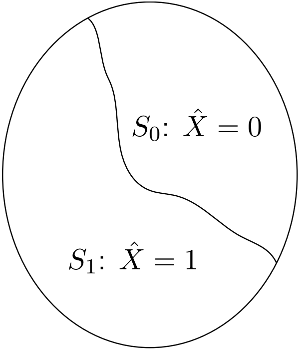


- $\mathcal{X}$: input alphabet
- $\mathcal{Y}$: output alphabet

Given $Y = y$ and Q, detect whether $X = 0$ or $X = 1$

Inference of Information

- given $Y = y$ and Q, detect whether $X = 0$ or $X = 1$
- **two types of error**: **false alarm** and **missed detection**

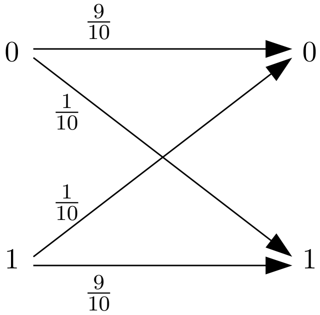


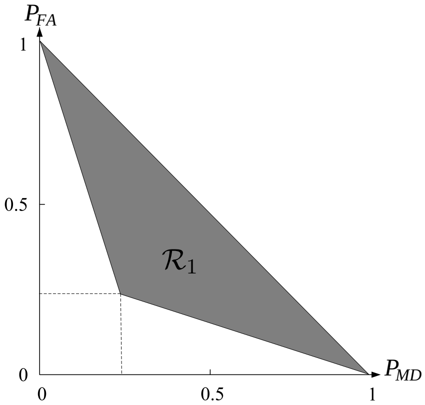
$\mathcal{Y}$: output alphabet

$P_{\mathrm{FA}} = \mathbb{P}\left(Y \in S_1 | X = 0\right)$ and $P_{\mathrm{MD}} = \mathbb{P}\left(Y \in S_0 | X = 1\right)$

Inference of Information

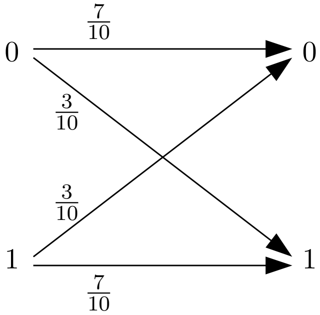
Case 1: Q_1

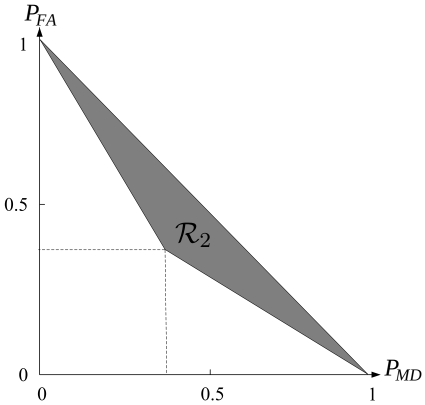




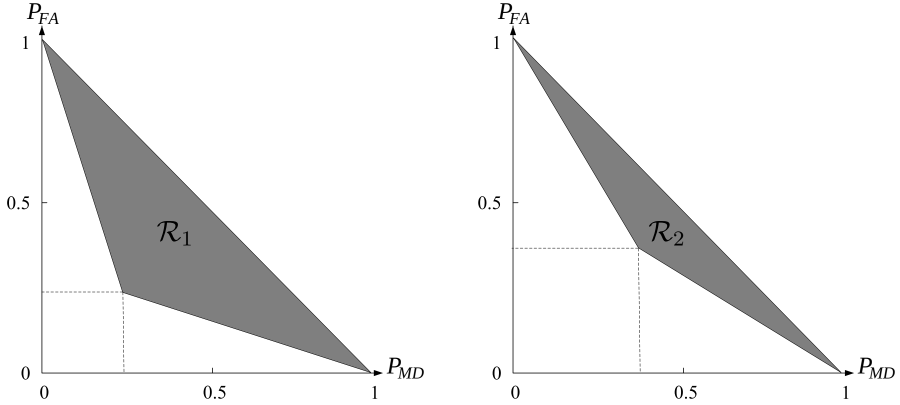
Inference of Information

Case 2: Q_2





Inference of Information



if $\mathcal{R}_2 \subset \mathcal{R}_1$, Q_2 guarantees **more privacy**

Local Differential Privacy

Q is ε-locally differentially private iff for all $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$e^{-\varepsilon} \leq \frac{Q(y|x)}{Q(y|x')} \leq e^{\varepsilon}$$

ε **controls the level of privacy**

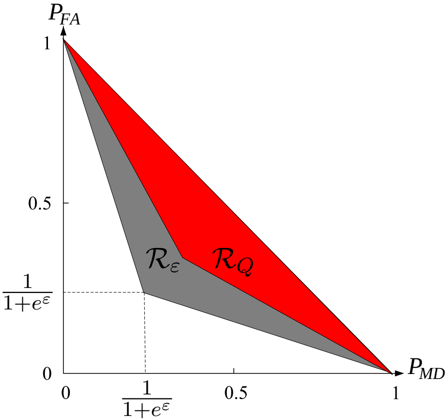
$\varepsilon \downarrow \implies$ **more private**

$\varepsilon \uparrow \implies$ **less private**

Local Differential Privacy

Q is ε-locally differentially private iff for all $x, x' \in \mathcal{X}$

$$\begin{aligned} P_{\text{FA}} + e^{\varepsilon} P_{\text{MD}} &\geq 1 \\ e^{\varepsilon} P_{\text{FA}} + P_{\text{MD}} &\geq 1 \end{aligned}$$



Qis ε-DP iff $\mathcal{R}_Q \subseteq \mathcal{R}_{\varepsilon}$ for all $x, x' \in \mathcal{X}$

Privacy vs. Utility

- the **more** private you want to be, the **less** utility you get
- there is a **fundamental trade-off** between **privacy** and **utility**

$$\begin{aligned} \underset{Q}{\text{maximize}} \quad & U(Q) \\ \text{subject to} \quad & Q \in \mathcal{D}_\varepsilon \end{aligned}$$

$U(Q)$: application dependent utility function

$\mathcal{D}_\varepsilon$: set of all ε-locally differentially private mechanisms

Summary of Results

Binary data: $|\mathcal{X}| = 2$

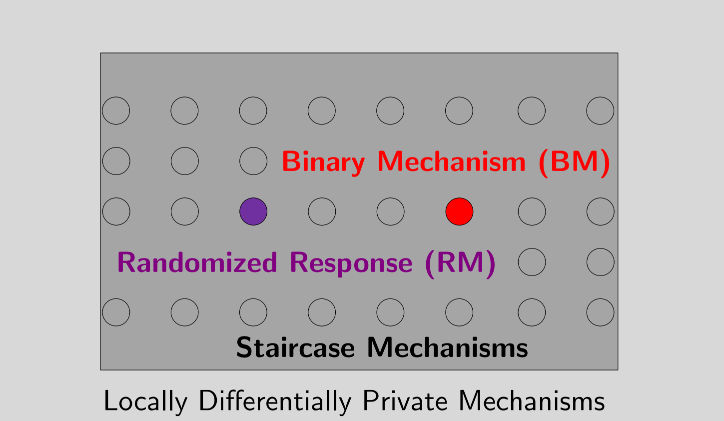
The Binary Randomized Response

w.p. $\frac{1}{1+e^{\varepsilon}}$ **lie** w.p. $\frac{e^{\varepsilon}}{1+e^{\varepsilon}}$ answer **truthfully**

- optimal for **all** ε
- optimal for **all** $U(Q)$ obeying the data processing inequality

Summary of Results

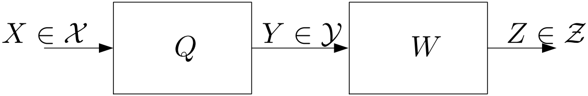
k-ary data: $|\mathcal{X}| = k > 2$



- staircase mechanisms are optimal for **all** ε and a **rich class of utilities**
- **BM** and **RR** are optimal in the **high** and **low** privacy regimes

CASE 1: BINARY DATA

Utility Functions

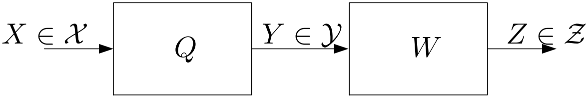


Utility functions obeying the data processing inequality:

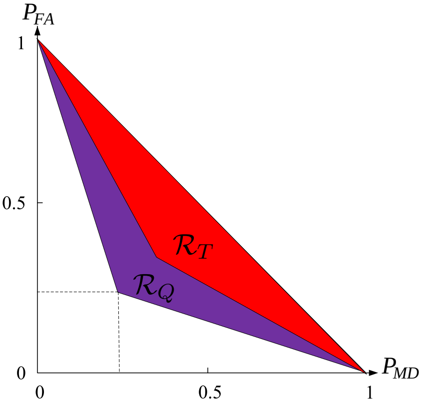
$$T = Q \circ W \implies U(T) \leq U(Q)$$

- further randomization can only reduce utility
- note that $Q \in \mathcal{D}_\varepsilon \implies T \in \mathcal{D}_\varepsilon$

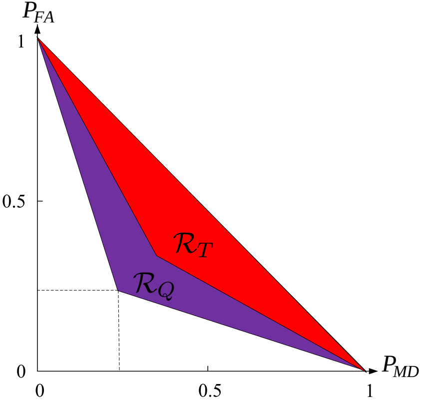
Data Processing Inequality (DPI)



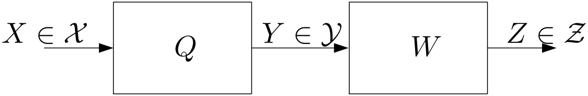
$$T = Q \circ W \implies \mathcal{R}_T \subseteq \mathcal{R}_Q$$



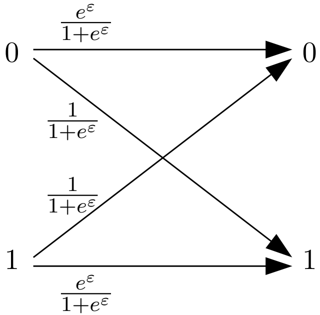
Converse to DPI [Blackwell 1953]

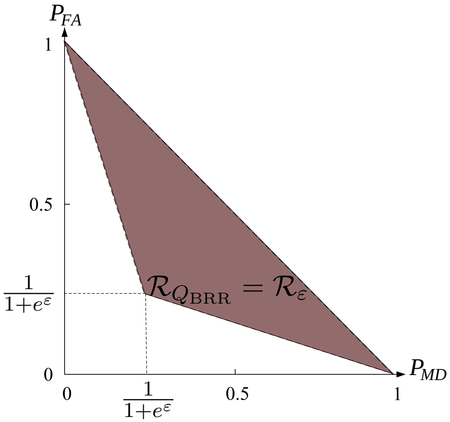


$$\mathcal{R}_T \subseteq \mathcal{R}_Q \implies \exists\, W \text{ s.t. } T = Q \circ W$$



Main Result [Kairouz, et. al., 2014]





$\forall \varepsilon, \forall Q \in \mathcal{D}_\varepsilon$: $\mathcal{R}_Q \subseteq \mathcal{R}_{Q_{\mathrm{BRR}}} \implies \exists\ W$ s.t. $Q = Q_{\mathrm{BRR}} \circ W$

$\implies \forall U$ obeying the data processing inequality: $U(Q) \leq U(Q_{\mathrm{RR}})$

The **binary randomized response** is **optimal**

CASE 2: k-ARY DATA

Information Theoretic Utility Functions

- $|\mathcal{X}| = k > 2$
- we focus on a rich class of convex functions

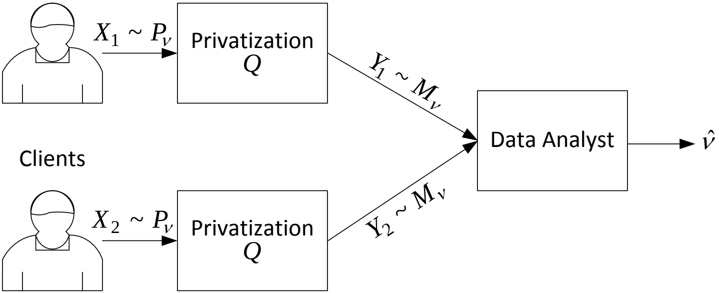
$$\begin{aligned} \underset{Q}{\text{maximize}} \quad & U(Q) = \sum_{y \in \mathcal{Y}} \mu(Q_y) \\ \text{subject to} \quad & Q \in \mathcal{D}_\varepsilon \end{aligned}$$

Q_y: the column of Q corresponding to $Q(y|\cdot)$

μ: any sublinear function

Includes all **f-divergences**, **mutual information**, etc.

Statistical Data Model



Analyst interested in **statistics** of data rather than **individual samples**

- X_i's are independently sampled from P_ν, $\nu \in \Lambda$
- **privatized data**: $Y_i \sim M_\nu = P_\nu \circ Q$

f-Divergences

For some convex function f such that $f(1) = 0$:

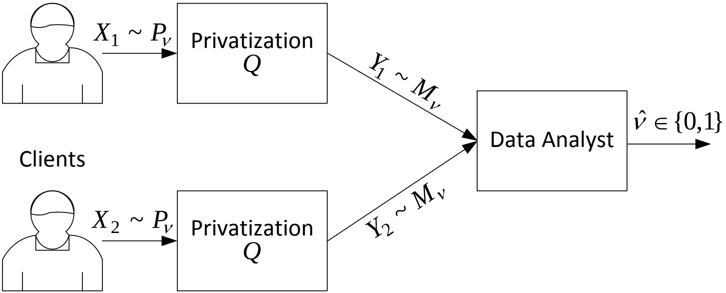
$$\begin{aligned} D_f(M_0 || M_1) &= \sum_{\mathcal{Y}} (P_1^T Q_y) f(P_0^T Q_y / P_1^T Q_y) \\ &= \sum_{\mathcal{Y}} \mu\left(Q_y\right) \end{aligned}$$

$P_\nu^T Q_y = \sum_{\mathcal{X}} Q(y|x) P_\nu(x)$
$\mu\left(Q_y\right) = (P_1^T Q_y) f(P_0^T Q_y / P_1^T Q_y)$

- KL divergence $D_{\mathrm{kl}}(M_0 || M_1)$
- total variation $\|M_0 - M_1\|_{\mathrm{TV}}$
- **minimax rates** and **error exponents**

Binary Hypothesis Testing

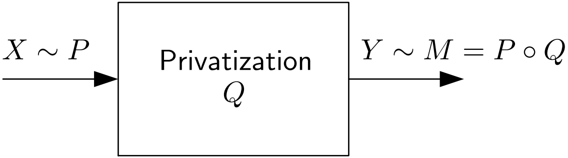
- n data providers: user i owns $X_i \in \mathcal{X}$
- X_i's are independently sampled from P_ν, $\nu \in \{\mathbf{0}, \mathbf{1}\}$



Given $\{Y_i\}_{i=1}^n$, detect whether $\nu = 0$ or $\nu = 1$

- Chernoff-Stein's lemma: $P_{\text{FA}} \approx e^{-n\, D_{\text{kl}}(M_0||M_1)}$
- for sufficiently small ε, $D_{\text{kl}}(M_0||M_1) \approx \varepsilon^2 D_{\text{kl}}(P_0||P_1)$

Information Preservation



Mutual Information between X and Y:

$$\begin{aligned} I(X;Y) &= \sum_{\mathcal{Y}} \sum_{\mathcal{X}} P(x)\, Q(y|x) \log \left(\frac{Q(y|x)}{\sum_{\mathcal{X}} P(x)\, Q(y|x)} \right) \\ &= \sum_{\mathcal{Y}} \mu(Q_y) \end{aligned}$$

$\mu(Q_y) = \sum_{\mathcal{X}} P(X=x)\, Q(y|x) \log \left(\frac{Q(y|x)}{\sum_{\mathcal{X}} P(x) Q(y|x)} \right)$

- for small ε, $I(X;Y) \approx \frac{1}{2} \max_{S \subseteq \mathcal{X}} \{P(S)P(S^c)\} \varepsilon^2$

Staircase Mechanisms

Recall that:

Q is ε-locally differentially private iff for all $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$

$$e^{-\varepsilon} \leq \frac{Q(y|x)}{Q(y|x')} \leq e^{\varepsilon}$$

ε **controls the level of privacy**

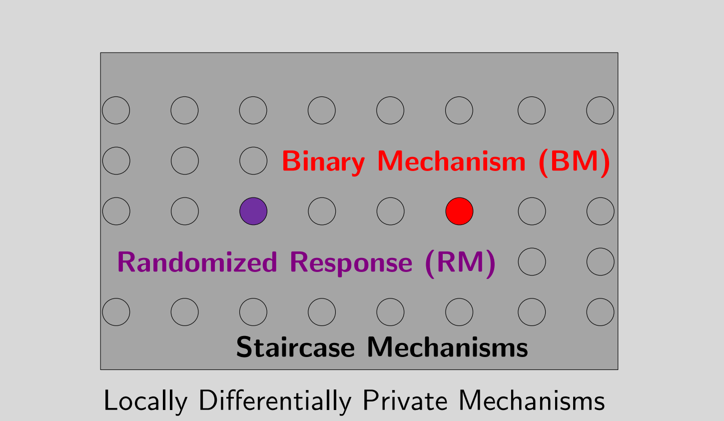
$\varepsilon \downarrow \implies$ **more private**

$\varepsilon \uparrow \implies$ **less private**

Staircase Mechanisms

Q is a **staircase mechanism** if for all $x, x' \in \mathcal{X}$ and $y \in \mathcal{Y}$:

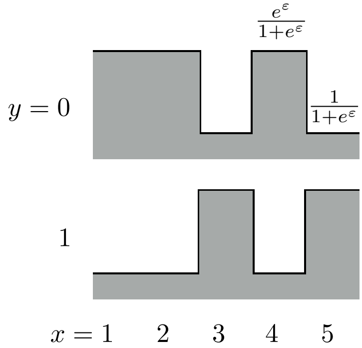
$$\frac{Q(y|x)}{Q(y|x')} \in \{e^{-\varepsilon}, 1, e^{\varepsilon}\}$$



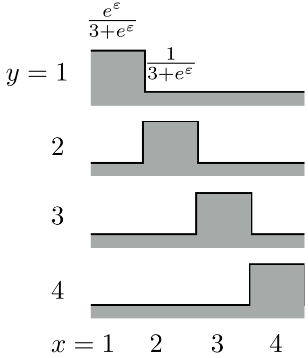
Examples of Staircase Mechanisms

$$Q^T = \frac{1}{1+e^{\varepsilon}} \begin{bmatrix} e^{\varepsilon} & e^{\varepsilon} & 1 & e^{\varepsilon} & 1 \\ 1 & 1 & e^{\varepsilon} & 1 & e^{\varepsilon} \end{bmatrix}$$

$$Q^T = \frac{1}{3+e^{\varepsilon}} \begin{bmatrix} e^{\varepsilon} & 1 & 1 & 1 \\ 1 & e^{\varepsilon} & 1 & 1 \\ 1 & 1 & e^{\varepsilon} & 1 \\ 1 & 1 & 1 & e^{\varepsilon} \end{bmatrix}$$



Binary Mechanism



Randomized Response

Main Results [Kairouz, et. al., 2014]

$\forall\ U(Q) = \sum_{y \in \mathcal{Y}} \mu(Q_y)$:

$$\begin{array}{llcll} \underset{Q}{\text{maximize}} & U(Q) & = & \underset{Q}{\text{maximize}} & U(Q) \\ \text{subject to} & Q \in \mathcal{D}_\varepsilon & & \text{subject to} & Q \in \mathcal{S}_\varepsilon \end{array}$$

$\mathcal{S}_\varepsilon$: set of all **staircase mechanisms** with $|\mathcal{Y}| \leq |\mathcal{X}|$

- **staircase** mechanisms are **optimal**
- no gain in **larger output alphabets**
- there are **finitely many** staircase mechanisms

For a given U, how do we find the optimal staircase mechanism?

Main Results [Kairouz, et. al., 2014]

$\forall\ U(Q) = \sum_{y \in \mathcal{Y}} \mu(Q_y)$:

$$\begin{array}{llcll} \underset{Q}{\text{maximize}} & U(Q) & = & \underset{\theta \in \mathbb{R}^{2^k}}{\text{maximize}} & \mu^T \theta \\ \text{subject to} & Q \in \mathcal{S}_\varepsilon & & \text{subject to} & S^{(k)}\theta = \mathbb{1} \\ & & & & \theta \geq 0 \end{array}$$

μ : 2^k-dimensional vector with $\mu_i = \mu(S_i^{(k)})$

$$S^{(3)} = \begin{bmatrix} 1 & 1 & 1 & 1 & e^\varepsilon & e^\varepsilon & e^\varepsilon & e^\varepsilon \\ 1 & 1 & e^\varepsilon & e^\varepsilon & 1 & 1 & e^\varepsilon & e^\varepsilon \\ 1 & e^\varepsilon & 1 & e^\varepsilon & 1 & e^\varepsilon & 1 & e^\varepsilon \end{bmatrix}$$

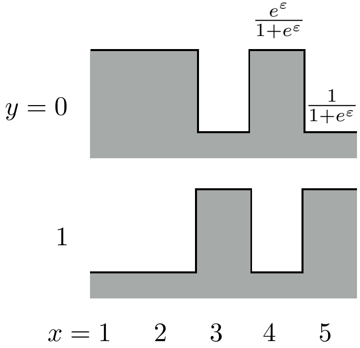
Main Results [Kairouz, et. al., 2014]

$\forall\ U(Q) = \sum_{y \in \mathcal{Y}} \mu(Q_y)$:

$$\begin{array}{llcll} \underset{Q}{\text{maximize}} & U(Q) & = & \underset{\theta \in \mathbb{R}^{2^k}}{\text{maximize}} & \mu^T \theta \\ \text{subject to} & Q \in \mathcal{S}_\varepsilon & & \text{subject to} & S^{(k)}\theta = \mathbb{1} \\ & & & & \theta \geq 0 \end{array}$$

- **finite dimensional linear program** of size 2^k
- **computationally expensive** if k is large
- do we really need to solve the problem?

Binary Mechanisms

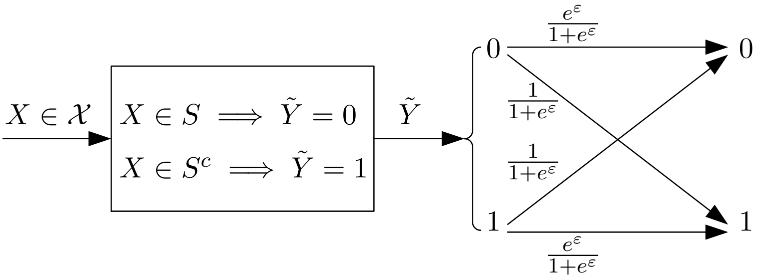


- maps k-**ary inputs** to **binary outputs**

Binary Mechanisms

A **deterministic** binary mapping followed by a **randomized response**

$\forall\ S \subseteq \mathcal{X}$:



- a highly quantized version of the original data

Optimality of Binary Mechanisms

f-divergences:

$$Q_{\mathrm{B}}(0|x) = \begin{cases} \frac{e^{\varepsilon}}{1+e^{\varepsilon}} & \text{if } P_0(x) \geq P_1(x) \\ \frac{1}{1+e^{\varepsilon}} & \text{if } P_0(x) < P_1(x) \end{cases}$$

$$Q_{\mathrm{B}}(1|x) = \begin{cases} \frac{e^{\varepsilon}}{1+e^{\varepsilon}} & \text{if } P_0(x) < P_1(x) \\ \frac{1}{1+e^{\varepsilon}} & \text{if } P_0(x) \geq P_1(x) \end{cases}$$

$\forall\ P_0, P_1$, $\exists\ \underline{\varepsilon}(P_0, P_1) > 0$ such that $\forall \varepsilon \leq \underline{\varepsilon}(P_0, P_1)$, Q_{B} is **optimal**

- $\forall \varepsilon$, Q_{B} is **optimal** for total variation distances

Optimality of Binary Mechanisms

Mutual Information:

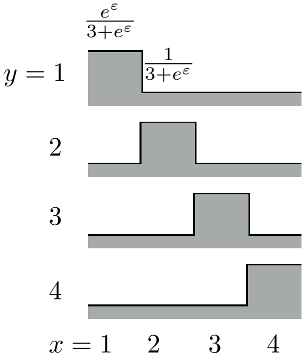
$$S^* \in \arg\max_{S \subseteq \mathcal{X}} P(S)P(S^c)$$

$$Q_{\mathrm{B}}(0|x) = \begin{cases} \frac{e^\varepsilon}{1+e^\varepsilon} & \text{if } x \in S^* \\ \frac{1}{1+e^\varepsilon} & \text{if } x \notin S^* \end{cases}$$

$$Q_{\mathrm{B}}(1|x) = \begin{cases} \frac{e^\varepsilon}{1+e^\varepsilon} & \text{if } x \notin S^* \\ \frac{1}{1+e^\varepsilon} & \text{if } x \in S^* \end{cases}$$

$\forall P$, $\exists\ \underline{\varepsilon}(P) > 0$ such that $\forall \varepsilon \leq \underline{\varepsilon}(P)$, Q_{B} is **optimal**

Randomized Response



- maps k-**ary inputs** to k-**ary outputs**

Randomized Response

w.p. $\frac{|\mathcal{X}|-1}{|\mathcal{X}|-1+e^{\varepsilon}}$ **lie** w.p. $\frac{e^{\varepsilon}}{|\mathcal{X}|-1+e^{\varepsilon}}$ answer **truthfully**

- **lie** $=$ choose another character in $\mathcal{X}$ uniformly at random
- can be viewed as a k-ary extension to the binary randomized response

Optimality of Randomized Response

KL Divergence:

$$Q_{\mathrm{RR}}(y|x) = \begin{cases} \frac{e^{\varepsilon}}{|\mathcal{X}|-1+e^{\varepsilon}} & \text{if } y = x \\ \frac{1}{|\mathcal{X}|-1+e^{\varepsilon}} & \text{if } y \neq x \end{cases}$$

$\forall\ P_0, P_1$, $\exists\ \overline{\varepsilon}(P_0, P_1) > 0$ such that $\forall \varepsilon \geq \overline{\varepsilon}(P_0, P_1)$, Q_{RR} is **optimal**

- note that Q_{RR} does not depend on P_0 and P_1

Optimality of Randomized Response

Mutual Information:

$$Q_{\mathrm{RR}}(y|x) = \begin{cases} \frac{e^{\varepsilon}}{|\mathcal{X}|-1+e^{\varepsilon}} & \text{if } y = x \\ \frac{1}{|\mathcal{X}|-1+e^{\varepsilon}} & \text{if } y \neq x \end{cases}$$

$\forall P$, $\exists\, \overline{\varepsilon}(P) > 0$ such that $\forall \varepsilon \geq \overline{\varepsilon}(P)$, Q_{RR} is **optimal**

- note that Q_{RR} does not depend on P

Big Picture

- local differential privacy is **crucial** for data collection applications
- we studied a broad class of information theoretic utilities
- we provided **explicit constructions** of **optimal mechanisms**

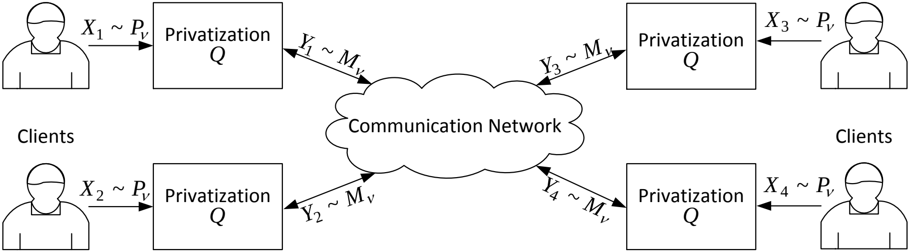
Big Picture

- local differential privacy is **crucial** for data collection applications
- we studied a broad class of information theoretic utilities
- we provided **explicit constructions** of **optimal mechanisms**

Big Picture

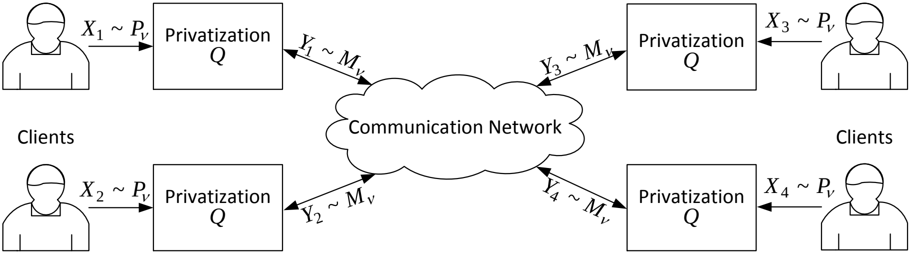
- local differential privacy is **crucial** for data collection applications
- we studied a broad class of information theoretic utilities
- we provided **explicit constructions** of **optimal mechanisms**

Private Multiparty Computation (PMC)



"Differentially Private Multi-party Computation: Optimality of
Non-Interactive Randomized Response
Peter Kairouz, Sewoong Oh, and Pramod Viswanath, 2014"

PMC: Main Results



- for binary data: use the simple binary randomized response
- **no cooperation needed!**
- for k-ary data: problem unsolved

Going Forward

- private **green button**
- private **genome sharing app**
- private **Google chrome** (RAPPOR)

Thank You
Questions?