How Well Does Privacy Compose?

Thomas Steinke

IBM Almaden

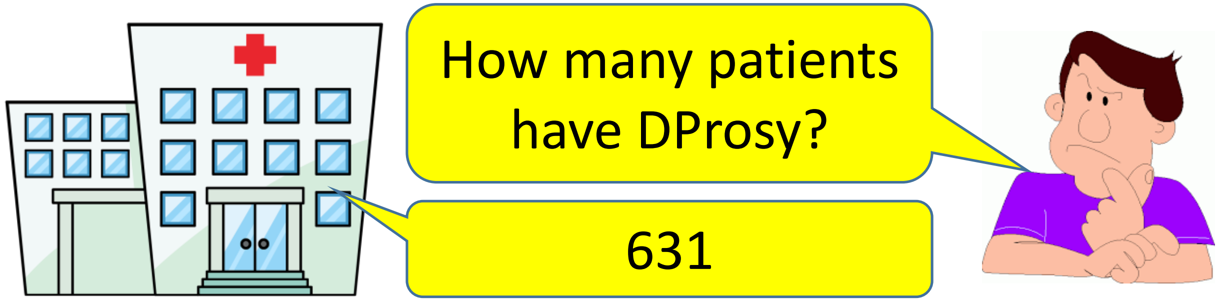
IPAM/UCLA, Los Angeles CA, 10 Jan. 2018

This Talk

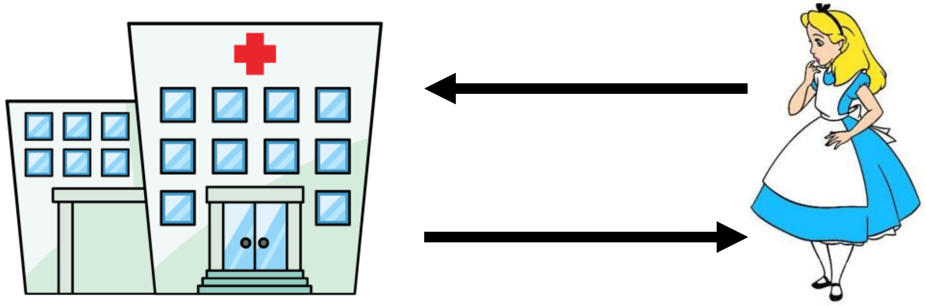
- Composition!!
  - What is composition?
  - Why is it important?
  - Composition & high-dimensional (e.g. genetic) data
- Concentrated differential privacy
  - Reformulation of DP with tight composition
  - Understand & compare to (ε, δ)-DP
  - Useful analytical tool & valuable theoretical perspective

What is composition?

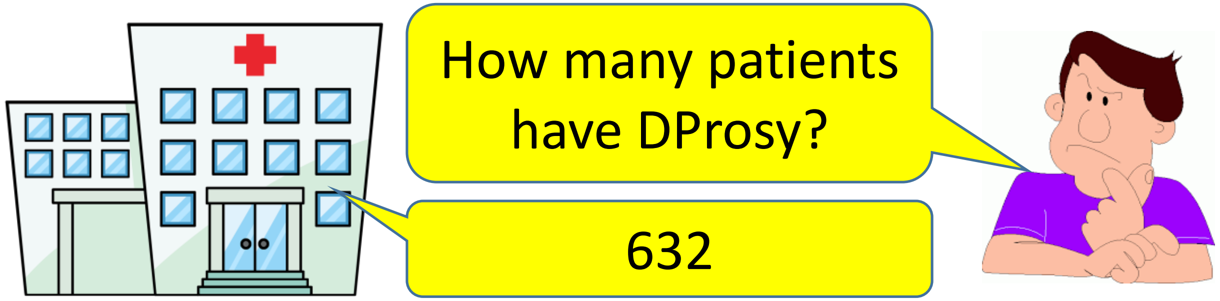
1:00 PM:



2:00 PM:



3:00 PM:



Conclusion: Alice was diagnosed with DProsy!

Here composition led to a privacy compromise.

Fortunately, DP protects against attacks like this.

Why is composition important?

- Your data is held by held by many entities who do not coordinate on privacy.
  Information released by these entities can be combined to violate privacy.

- Allows complex algorithms to be built -- crucial for handling high-dimensional data (e.g. genetic data).

High-dimensional data & one-way marginals

Dimension d

Alice	0	0	1	0	0	1	1	1	0	0	1
Bob	1	0	1	1	0	1	1	0	0	1	1
Charles	1	0	1	0	1	1	0	0	0	0	0
David	0	1	0	1	1	1	0	1	0	0	1

#individuals n

- E.g. GWAS data. $d \approx 10^6, n \approx 1000$
- **Key Question: For a given n and d, how accurately can we release the one-way marginals of this dataset without imperiling privacy?**
- I.e. how does privacy risk compose over the attributes?

Privacy risks of one-way marginals

Dimension d (columns); #individuals n (rows)

Alice	0	0	1	0	0	1	1	1	0	0	1
Bob	1	0	1	1	0	1	1	0	0	1	1
Charles	1	0	1	0	1	1	0	0	0	0	0
David	0	1	0	1	1	1	0	1	0	0	1
	.5	.25	.75	.5	.5	1	.5	.5	0	.25	.75

- [Homer+08, Sankararaman+09, Bun+14, Dwork+15, etc.] showed that one-way marginals are susceptible to <u>tracing</u>.
- That is, given someone's data and the one-way marginals of a case group, we can determine whether that person is in the group.
  - Surprising!
  - Led to privacy policy changes by NIH.
  - Works as long as $d \gg n$.
  - Works even with approximate one-way marginals (but requires larger d).

This Talk

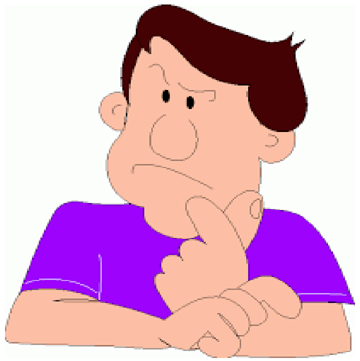
- Composition!!
  - What is composition?
  - Why is it important?
  - Composition & high-dimensional (e.g. genetic) data
- Concentrated differential privacy
  - Reformulation of DP with tight composition
  - Understand & compare to (ε, δ)-DP
  - Useful analytical tool & valuable theoretical perspective

Differential Privacy [DMNS06…]

Alice	0	1	1	0	0	1	0	1
Bob	1	0	1	1	0	0	1	1
Charles	1	0	1	0	1	1	0	0
David	0	1	0	1	1	1	0	1

Sensitive Dataset

Interface



Analyst

Definition: A randomized algorithm M is **differentially private** if, for all datasets x and x' differing only on one individual's data,

$$\mathrm{distribution}\big(M(x)\big) \approx \mathrm{distribution}\big(M(x')\big).$$

Noisy one-way marginals

Dimension d

#individuals n

Alice	0	0	1	0	0	1	1	1	0	0	1
Bob	1	0	1	1	0	1	1	0	0	1	1
Charles	1	0	1	0	1	1	0	0	0	0	0
David	0	1	0	1	1	1	0	1	0	0	1
One-way marginals	.5	.25	.75	.5	.5	1	.5	.5	0	.25	.75
Noisy marginals	.6	.1	.8	.6	.4	1	.4	.5	.1	.4	.9

Adding normally-distributed noise to all the values satisfies DP.

Does this give good privacy-utility tradeoff?

Quantifying Differential Privacy

Rényi divergence [R61]:

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1}\log\left(\int_\Omega P(x)^\alpha Q(x)^{1-\alpha}\, dx\right)$$

Interpolates between KL divergence ($\alpha \to 1$) & max divergence ($\alpha \to \infty$).

Exactly characterizes adding Normal noise.

- ε-DP [DMNS06]:

$$\forall y \qquad \mathbb{P}[M(x) = y] \;\leq\; e^{\varepsilon} \ldots y]$$

- (ε, δ)-DP [DKMMN06]:

$$\forall S \qquad \mathbb{P}[M(x) \in S] \;\leq\; e^{\varepsilon} \ldots (x') \in S] + \delta$$

New! ρ-CDP [DR16,BS16,M17,BDRS17]:

$$\forall \alpha \in (1, \infty) \qquad D_\alpha(M(x)||M(x')) \leq \rho\alpha$$

Why do we need a new definition?

- "Pure" ε-DP gives poor composition bounds
  - Gets "hung up on" very low probability events.
  - Composition is quadratically worse than it "should" be.
- "Approximate" (ε, δ)-DP gives messy composition bounds
  - Can ignore events with probability $\leq \delta$.
  - Doesn't sharply capture what's going on.
  - Superfluous $\log(1/\delta)$ factors in composition analysis.
- Concentrated DP gives sharp composition bounds!

$\delta = \mathbb{P}[\text{bad event}]$ needs to be cryptographically small.

Composition & privacy loss are natural phenomena

Composition for CDP

<u>Theorem</u> (CDP composition [DR16,BS16]):
Let $M_1, \ldots, M_k$ be randomized algorithms. Suppose each M_i is ρ_i-CDP.
Then combining the outputs of $M_1, \ldots, M_k$ satisfies $(\rho' = \sum_i \rho_i)$-CDP.

- Simple and optimal (in contrast to ε-DP and (ε, δ)-DP).

- Cf. Optimal (ε, δ)-DP composition [KOV15,MV16]:

$$\frac{\sum_{s \subseteq [k]} \max\left\{0, e^{\sum_{i \in s} \varepsilon_i} - e^{\varepsilon' + \sum_{i \in [k] \setminus s} \varepsilon_i}\right\}}{\prod_{i \in [k]} (1 + e^{\varepsilon_i})} + \frac{1 - \delta'}{\prod_{i \in [k]} (1 - \delta_i)} \leq 1$$

  - Computing optimal composition exactly is #P-hard [MV16]!!

Noisy one-w

Given $q_1, \ldots, q_k : X \to [0,1]$ and private dataset $x \in X^n$ output $a_1, \ldots, a_k \in [0,1]$ such that with high probability

$$\frac{1}{k}\sum_{j=1}^{k}\left|a_j - \frac{1}{n}\sum_{i=1}^{n} q_j(x_i)\right| \le \frac{1}{100}$$

More sophisticated

E.g.,

- Only identify the k most sign... ...butes.
- Attributes are sparse/structured.
- Exploit data distribution.

One-way	.5	.25	.75	.5	.5	1	.5	.5	0	.25	.75
Noisy ma	.6	.1	.8	.6	.4	1	.4	.5	.1	.4	.9

Adding $N(0, \sigma^2)$ to each marginal achieves $\left(\rho = \frac{d}{2\sigma^2 n^2}\right)$-CDP.

Sharp tradeoff between privacy ρ, dimension d, accuracy σ, and number of individuals n.

e.g. $\rho = 0.5$, $\sigma = 0.1$, $d = 10^4$ requires $n = \frac{\sqrt{d}}{\sigma\sqrt{2\rho}} = 1000$.

Composition Comparison

- Pure ε-DP: $\varepsilon' = \sum_i \varepsilon_i$.
  - Can approximate $\mathrm{d} = \Theta(\varepsilon n)$ one-way marginals to constant accuracy with ε-DP.
- Approx. (ε, δ)-DP: $\varepsilon' = O\left(\sqrt{\log(1/\delta') \sum_{i=1}^{k} \varepsilon_i^2}\right), \delta' = O\left(\sum_{i=1}^{k} \delta_i\right)$.
  - #P-hard to compute optimal composition exactly.
  - Can approximate $\mathrm{d} = \Theta(\varepsilon^2 n^2 / \log(1/\delta))$ marginals to constant accuracy with (ε, δ)-DP.
- ρ-CDP: $\rho' = \sum_i \rho_i$.
  - Can approximate $d = \Theta(\rho n^2)$ marginals to constant accuracy with ρ-CDP.

Linear

Almost Quadratic

Quadratic

log factor "absorbed"

This Talk

- Composition!!
  - What is composition?
  - Why is it important?
  - Composition & high-dimensional (e.g. genetic) data
- Concentrated differential privacy
  - Reformulation of DP with tight composition
  - Understand & compare to (ε, δ)-DP
  - Useful analytical tool & valuable theoretical perspective

Concentrated DP [DR16,BS16, BDRS17]

Definition [BS16]: A randomized algorithm M is ρ-CDP if, for all datasets x and x' differing only on one individual's data,

$$\forall \alpha \in (1,\infty) \qquad D_\alpha(M(x)||M(x')) \leq \rho\alpha$$

Rényi divergence [R61]:

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1}\log\left(\int_\Omega P(x)^\alpha Q(x)^{1-\alpha}\, dx\right)$$

Interpolates between KL divergence ($\alpha \to 1$) & max divergence ($\alpha \to \infty$).

Exactly characterizes Gaussian mechanism.

CDP versus $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$-DP

Theorem [BS16]: $\forall \rho, \delta > 0$

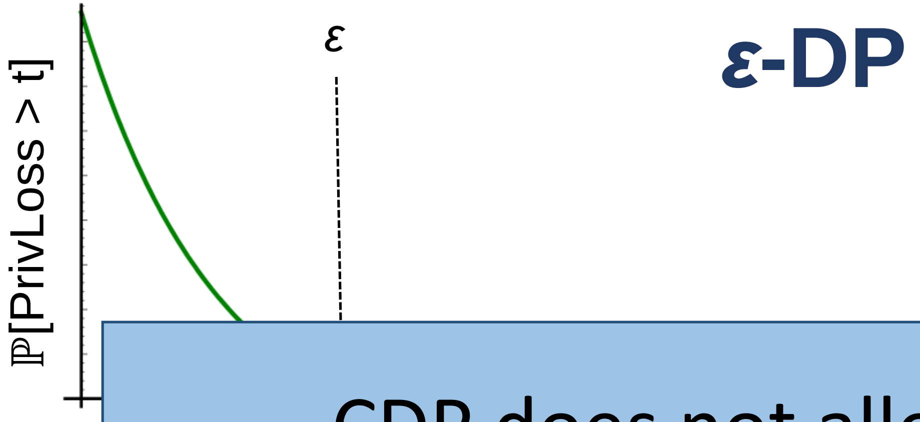
$$\sqrt{2\rho}\text{-DP} \implies \rho\text{-CDP} \implies \left(\rho + 2\sqrt{\rho \cdot \log(1/\delta)}, \delta\right)\text{-DP}.$$

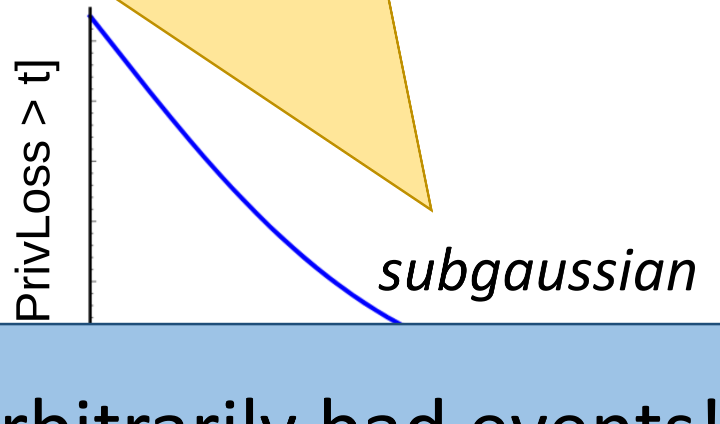
- CDP is a relaxation of pure ε-DP.
  - Relaxation is strict. E.g. Gaussian mechanism satisfies CDP, but not pure DP.
- CDP is roughly equivalent to approx. (ε, δ)-DP with this $\forall\delta$ quantification.
  - However, there are algorithms that satisfy approx. DP, but not CDP.
- Think of CDP as being intermediate between pure and approx. DP.
- Open-ended question: How to interpret ρ?

CDP versus $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$-D

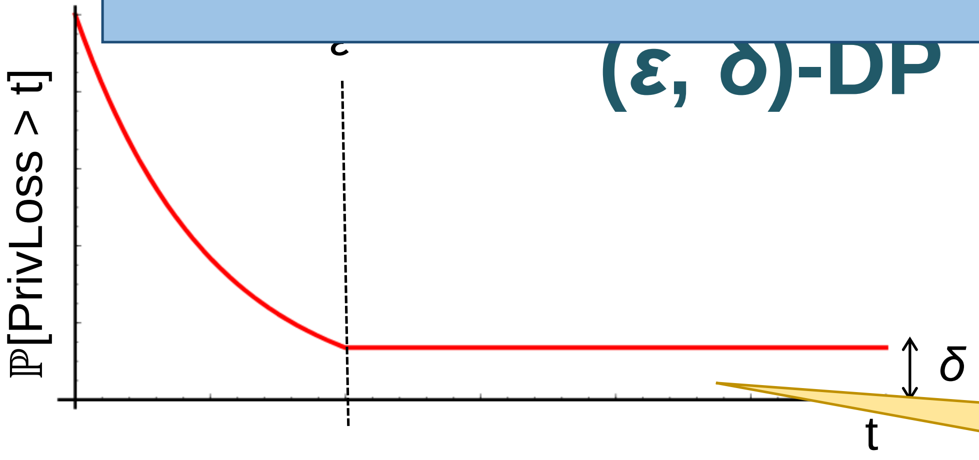
Smooth control of bad events: For ρ-CDP

$\forall t > \rho \qquad \mathbb{P}[\text{PrivLoss} > t] \leq e^{-(t-\rho)^2/4\rho}$





CDP does not allow arbitrarily bad events!



$$\text{PrivLoss}(y) = \log\left(\frac{\mathbb{P}[M(x) = y]}{\mathbb{P}[M(x') = y]}\right)$$

$$\text{PrivLoss} = \text{PrivLoss}(M(x))$$

With probability δ arbitrarily bad things can happen!

Bounding Bad Events with CDP [M17]

<u>Proposition</u> [M17]: If M is ρ-CDP and x, x' are neighbouring inputs, then

$$\forall S\ \forall \alpha \quad \mathbb{P}[M(x) \in S] \le e^{(\alpha-1)\rho} \cdot (\mathbb{P}[M(x') \in S])^{1-1/\alpha}$$

E.g.:

- Suppose, when not in dataset, bad event happens with
$$\mathbb{P}[M(x') \in S] \le 10^{-10}$$
- If M is ρ -CDP, then, when in data, bad event happens with
$$\alpha = 2: \quad \mathbb{P}[M(x) \in S] \le e^{\rho}\sqrt{10^{-10}}$$
$$\alpha = 10: \quad \mathbb{P}[M(x) \in S] \le e^{9\rho} 10^{-9}$$

This Talk

- Composition!!
  - What is composition?
  - Why is it important?
  - Composition & high-dimensional (e.g. genetic) data
- Concentrated differential privacy
  - Reformulation of DP with tight composition
  - Understand & compare to (ε, δ)-DP
  - Useful analytical tool & valuable theoretical perspective

What can we do with CDP?

(ε, δ)-DP

Propose-Test-Release framework,
Smooth sensitivity

ρ-CDP

Advanced composition,
Gaussian mechanism,
Private multiplicative weights,
Projection mechanism

ε-DP

Basic composition,
Laplace mechanism,
Exponential mechanism,
Randomized response,
Sparse vector,
BLR mechanism

Truncated CDP [BDRS17]

Definition [BDRS17]: A randomized algorithm M is (ρ, ω)-tCDP if, for all datasets x and x' differing only on one individual's data,

$$\forall \alpha \in (1, \omega) \qquad D_\alpha(M(x)||M(x')) \leq \rho\alpha$$

- $\omega = \infty$ recovers ρ-CDP.
- Similar to Rényi DP [M17] – consider single α, rather than interval.
- Extends CDP to permit analogs of key algorithmic techniques.
  - Analog of propose-test-release framework [DL09].
  - Smooth sensitivity [NRS07].
  - Privacy amplification by subsampling.

This Talk

- Composition!!
  - What is composition?
  - Why is it important?
  - Composition & high-dimensional (e.g. genetic) data
- Concentrated differential privacy
  - Reformulation of DP with tight composition
  - Understand & compare to (ε, δ)-DP
  - Useful analytical tool & valuable theoretical perspective

Separation: $\boldsymbol{\varepsilon}$-DP $\neq$ CDP $\neq$ $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$-DP

Point Queries/Histograms:

Input: $x_1, \dots, x_n \in \Omega$.

Output: For each $z \in \Omega$, return $\mathrm{freq}(z) = |\{i : x_i = z\}| \pm \frac{n}{100}$.

- Possible with ε-DP iff $n = \Theta(\log|\Omega| / \varepsilon)$.
- Possible with (ε, δ)-DP iff $n = \Theta(\log(1/\delta) / \varepsilon)$.
- Possible with ρ-CDP iff $n = \Theta\left(\sqrt{\log|\Omega| / \rho}\right)$.
  - Upper bound: Add noise from $\mathcal{N}(0, \frac{1}{\rho})$ to each frequency.

Quadratic separation

"Infinite" separation

CDP & Mutual Information

Theorem [BS16]: If M is ρ-CDP and X is a random input consisting of n individuals, then

$$I\big(X; M(X)\big) \leq \rho \cdot n^2$$

- Follows from group privacy property of CDP.
- Idea: If M is accurately answers many queries, then mutual information must be high.
- $\Longrightarrow$ Lower bound on n.

Separation: ε-DP $\neq$ CDP $\neq$ (ε, δ)-DP

Point Queries/Histograms:

Input: $x_1, \ldots, x_n \in \Omega$.

Output: For each $z \in \Omega$, return $\text{freq}(z) = |\{i : x_i = z\}| \pm \frac{n}{100}$.

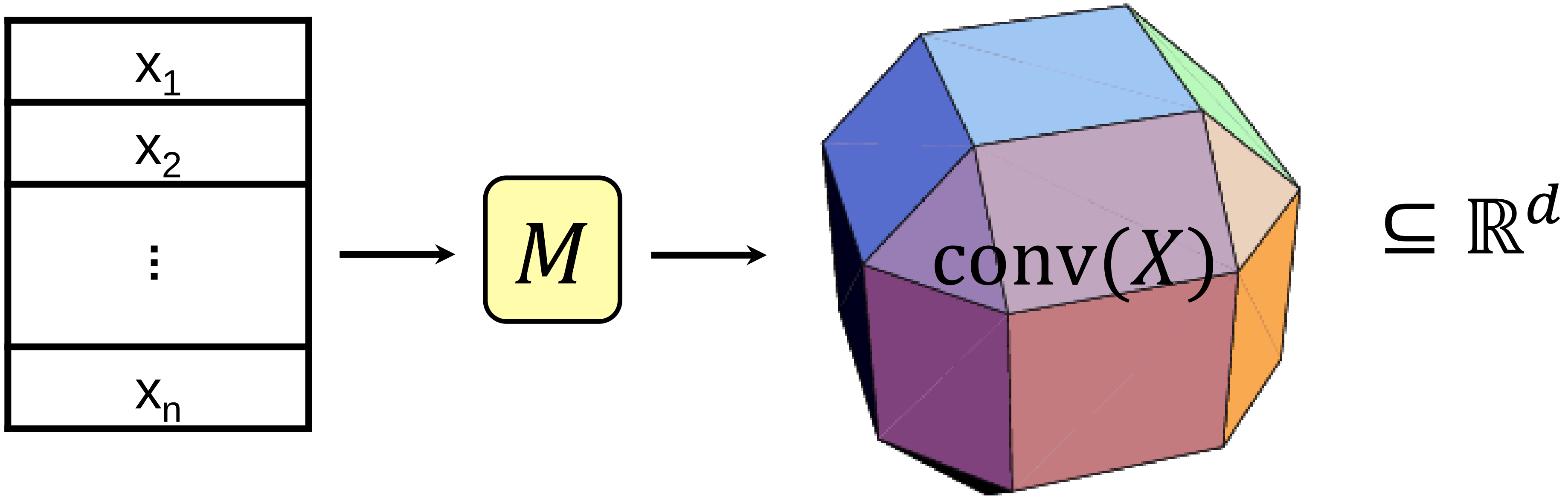
- Possible with ε-DP iff $n = \Theta(\log|\Omega| / \varepsilon)$.
- Possible with (ε, δ)-DP iff $n = \Theta(\log(1/\delta) / \varepsilon)$.
- Possible with ρ-CDP iff $n = \Theta\left(\sqrt{\log|\Omega| / \rho}\right)$.
  - Upper bound: Add noise from $\mathcal{N}(0, \frac{1}{\rho})$ to each frequency.

Quadratic separation

“Infinite” separation

Optimal CDP Algorithm [BBNS17] (see poster)

Linear Query Release problem:

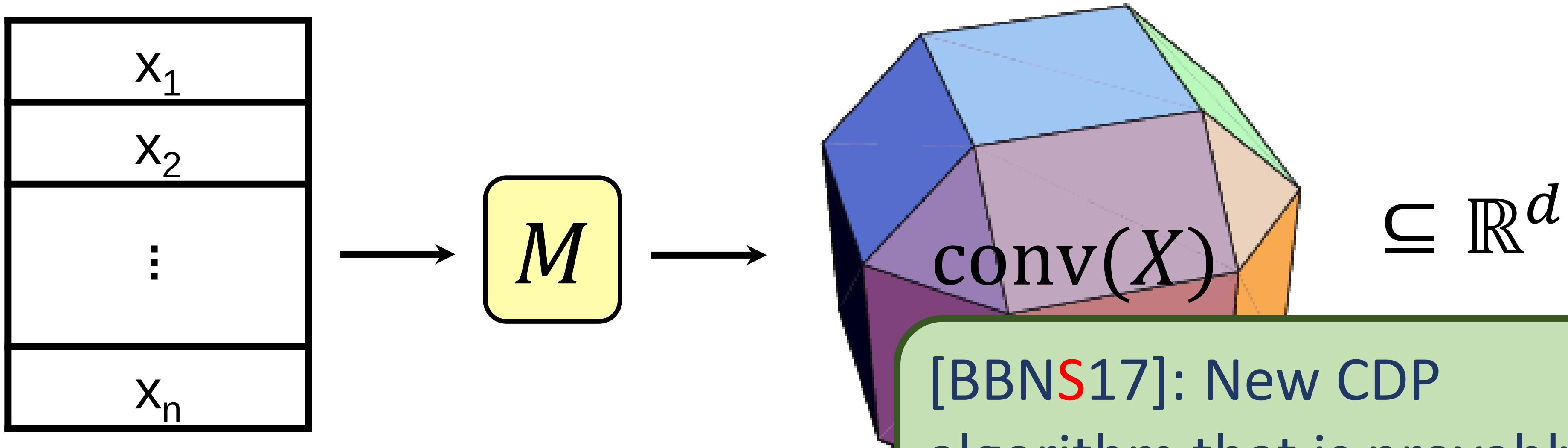


Accuracy Goal:

$$M(x) \approx_\alpha \frac{1}{n}\sum_{i=1}^{n} q(x_i)$$

Optimal CDP Algorithm [BBNS17] (see poster)

Linear Query Release problem:



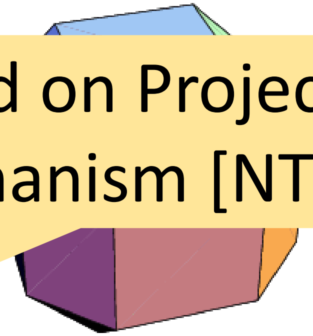
[BBNS17]: New CDP algorithm that is provably optimal for $\alpha = \Omega(1)$.

Average Squared Accuracy Goal:

$$\mathbb{E}\left[\frac{1}{\mathrm{diam}(X)^2}\left\|M(x) - \frac{1}{n}\sum_{i=1}^{n} q(x_i)\right\|_2^2\right] \leq \alpha^2$$

Optimal CDP Algorithm [BBNS17] (see poster)

Let $X = \text{range}(q) \subseteq \mathbb{R}^d$ be the set of possible answers



Based on Projection Mechanism [NTZ13]

Definition (Covering number): Let $N(X, \gamma)$ be the smallest number of γ-balls whose union covers X.

Algorithm [BBNS17]: ρ-CDP α-accurate algorithm for X as long as

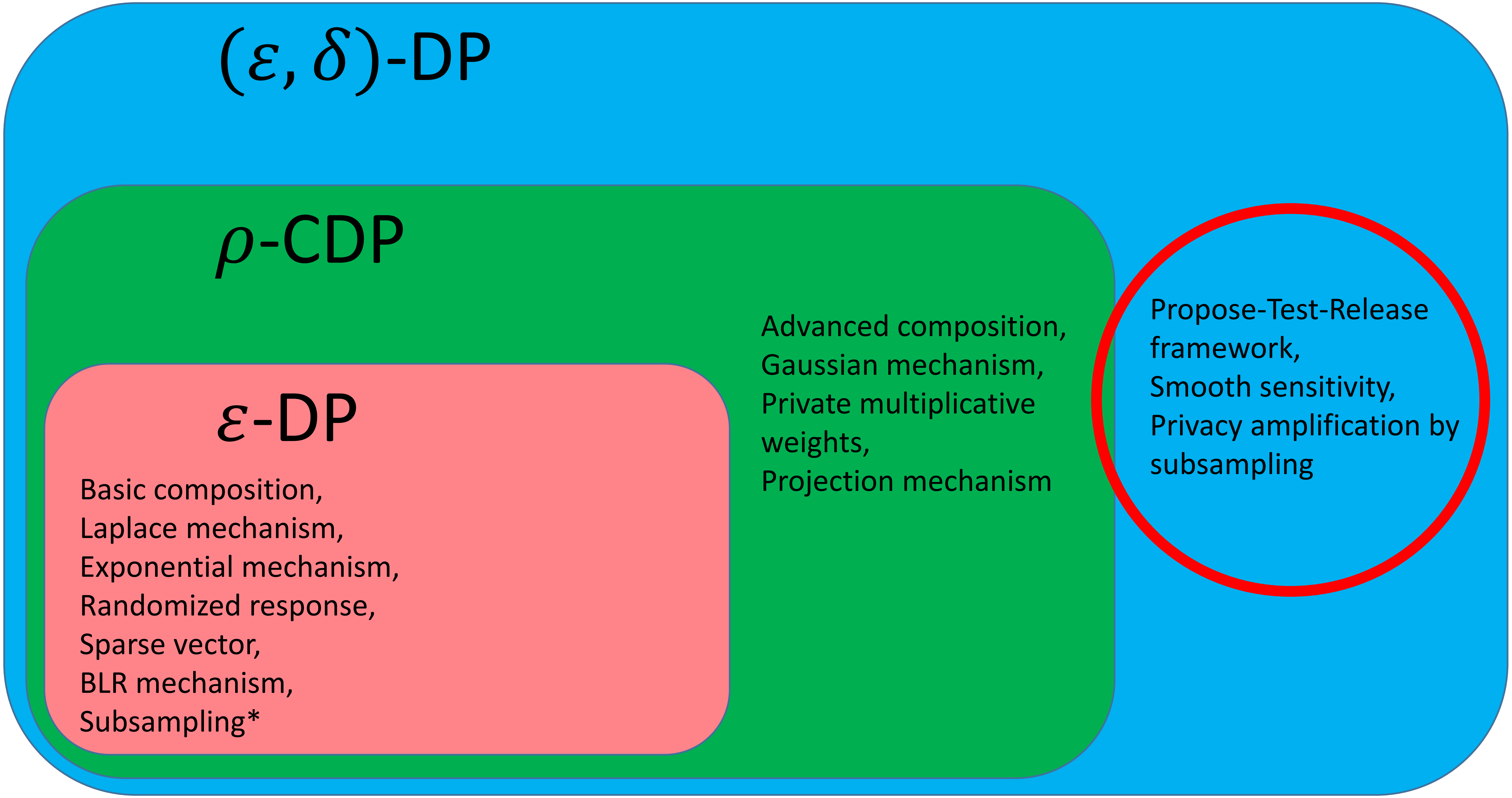
$$n \geq O\left(\frac{1}{\alpha^2}\sqrt{\frac{\log N(X, \alpha \cdot \text{diam}(X)/2)}{\rho}}\right)$$

Lower Bound [BBNS17]: Need $n \geq \Omega\left(\sqrt{\log N(X, 3\alpha \cdot \text{diam}(X))/\rho}\right)$.

Average Squared Accuracy Goal:

$$\mathbb{E}\left[\frac{1}{\text{diam}(X)^2}\left\|M(x) - \frac{1}{n}\sum_{i=1} q(x_i)\right\|_2\right] \leq \alpha^2$$

What can't we do with CDP?



Truncated CDP [BDRS17]

Definition [BDRS17]: A randomized algorithm M is (ρ, ω)-tCDP if, for all datasets x and x' differing only on one individual's data,

$$\forall \alpha \in (1, \omega) \qquad D_\alpha(M(x)||M(x')) \leq \rho\alpha$$

- $\omega = \infty$ recovers ρ-CDP.
- Similar to Rényi DP [M17] – consider single α, rather than interval.
- Extends CDP to permit analogs of key algorithmic techniques.
  - Propose-test-release framework.
  - Smooth sensitivity.
  - Privacy amplification by subsampling.

Separation: ε-DP $\neq$ CDP $\neq$ tCDP

<u>Point Queries/Histograms</u>:
Input: $x_1, \ldots, x_n \in \Omega$.
Output: For each $z \in \Omega$, return $\mathrm{freq}(z) = |\{i : x_i = z\}| \pm \frac{n}{100}$.

- Possible with ε-DP iff $n = \Theta(\log|\Omega|/\varepsilon)$.
- Possible with (ε, δ)-DP iff $n = \Theta(\log(1/\delta)/\varepsilon)$.
- Possible with ρ-CDP iff $n = \Theta\left(\sqrt{\log|\Omega|/\rho}\right)$.
- **Possible with (ρ, ω)-tCDP iff $n = \Theta(\omega \cdot \log\log|\Omega|)$ (for $\omega \ll \sqrt{\log|\Omega|/\rho}$).**

Related to propose-test-release.

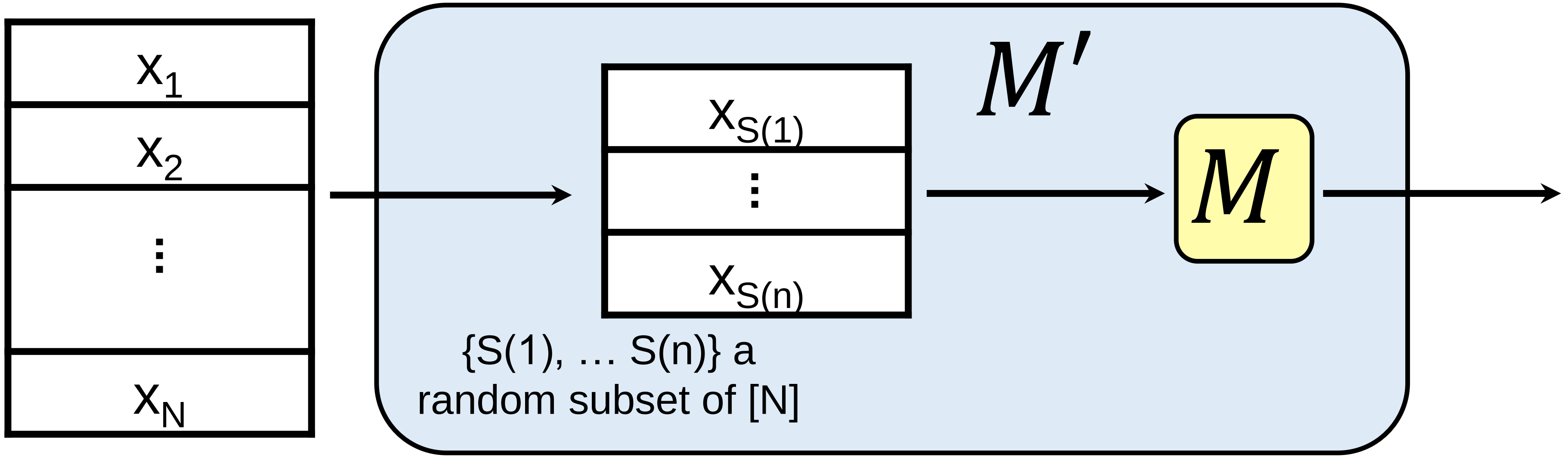
How do these compare in practice?

Subsampling

$M: X^n \to Y$

$M': X^N \to Y$

$s = \frac{n}{N} \ll 1$



$\log(1 + s \cdot (e^{\varepsilon} - 1))$

- If M is (ε, δ)-DP, then M' is $(\approx s \cdot \varepsilon, s \cdot \delta)$-DP.
- If M is ρ-CDP, then M' is ρ-CDP. No gain in parameters!
- **If M is (ρ, ω)-tCDP, then M' is $(\approx s^2 \cdot \rho, \Omega(\min\{\omega, \log(1/s)/\rho\}))$-tCDP.**

What can't we do with tCDP?

(ε, δ)-DP

(ρ, ω)-tCDP

???

ρ-CDP

ε-DP

Basic composition,
Laplace mechanism,
Exponential mechanism,
Randomized response,
Sparse vector,
BLR mechanism,
Subsampling*

Advanced composition,
Gaussian mechanism,
Private multiplicative weights,
Projection mechanism

Propose-Test-Release framework,
Smooth sensitivity,
Privacy amplification by subsampling