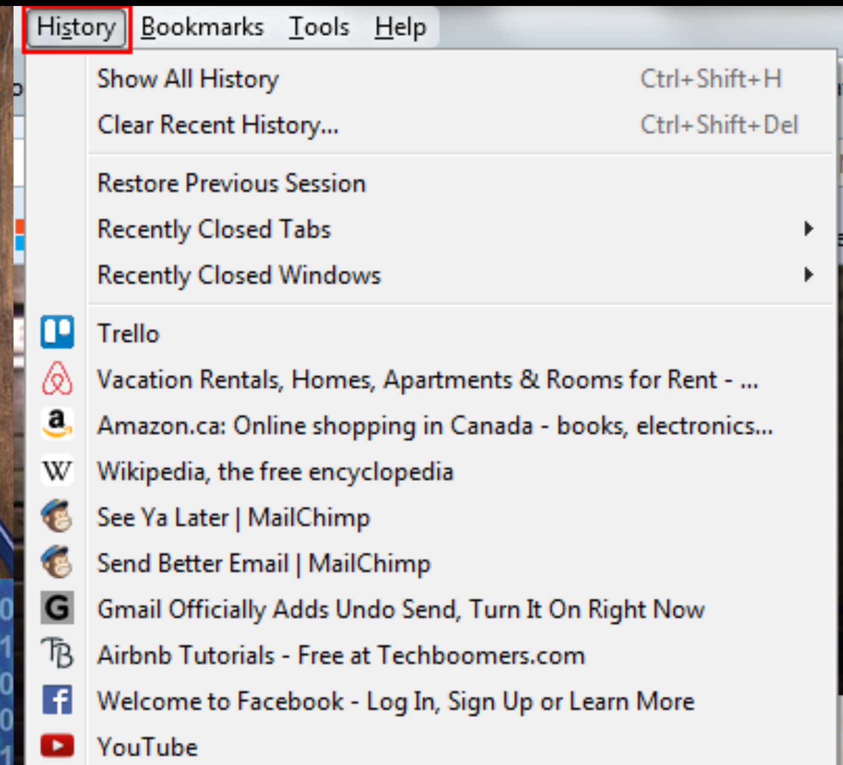
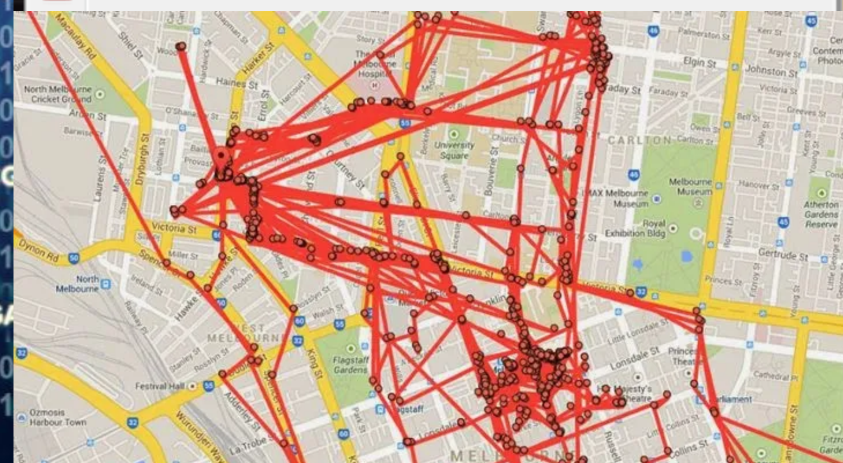
Privacy, Robustness, and Statistics in High Dimensions

Samuel B. Hopkins, MIT

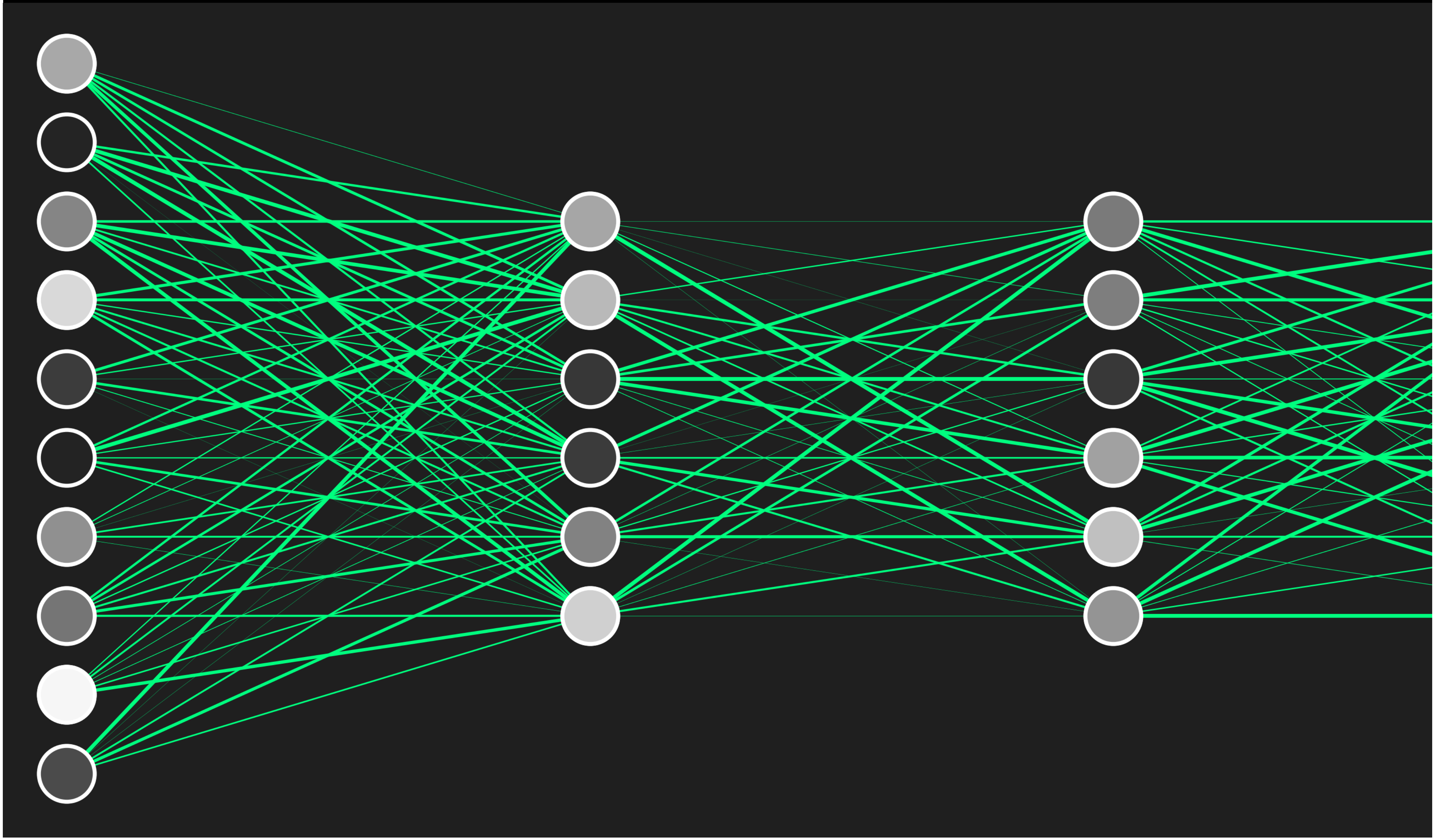
TTI Chicago, 2024


Medical record
Home
Service
Shedule
Laboratory
Patients


History
Bookmarks
Tools
Help
Show All History Ctrl+Shift+H
Clear Recent History... Ctrl+Shift+Del
Restore Previous Session
Recently Closed Tabs
Recently Closed Windows
Trello
Vacation Rentals, Homes, Apartments & Rooms for Rent - ...
Amazon.ca: Online shopping in Canada - books, electronics...
Wikipedia, the free encyclopedia
See Ya Later | MailChimp
Send Better Email | MailChimp
Gmail Officially Adds Undo Send, Turn It On Right Now
Airbnb Tutorials - Free at Techboomers.com
Welcome to Facebook - Log In, Sign Up or Learn More
YouTube



How can we utilize data for science, industry, medicine,…



…without violating individual privacy?

Differential Privacy (DP)

A is ε-DP if for every pair of inputs X, X' differing on one individual, and every output o,

$$\Pr(A(X) = o) \leq e^{\varepsilon} \cdot \Pr(A(X') = o) \ .$$

[DMNS06]

Differential Privacy (DP)

A is ε-DP if for every pair of inputs X, X' differing on one individual, and every output o,

$$\Pr(A(X) = o) \leq e^{\varepsilon} \cdot \Pr(A(X') = o) \ .$$

For most private statistics: one individual = one sample

[DMNS06]

Differential Privacy (DP)

Consequence: Hypothesis tests to distinguish $A(X)$, $A(X')$ have:

$$Type\ I\ error\ +\ Type\ II\ error \geq 1 - O(\varepsilon).$$

Approximate Differential Privacy

A is (ε, δ)-DP if for every pair of inputs X, X' differing on one individual, and every event E,

$$\Pr(A(X) \in E) \leq e^{\varepsilon} \cdot \Pr(A(X') \in E) + \delta.$$

[DMNS06]

Approximate Differential Privacy

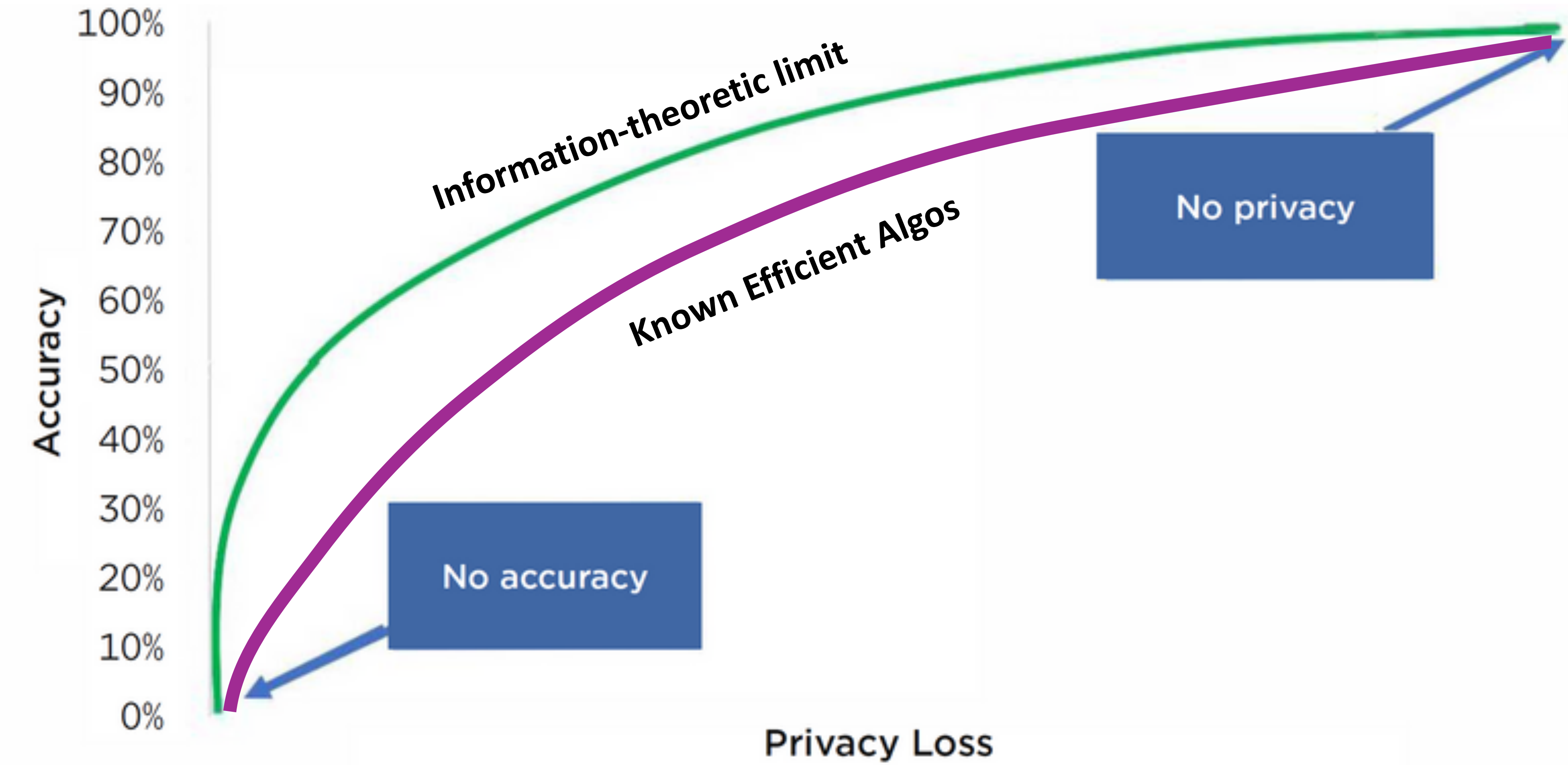
A is (ε, δ)-DP if for every pair of inputs X, X' differing on one individual, and every event E,

$$\Pr(A(X) \in E) \leq e^{\varepsilon} \cdot \Pr(A(X') \in E) + \delta.$$

Other variants: concentrated DP, Renyi DP,…

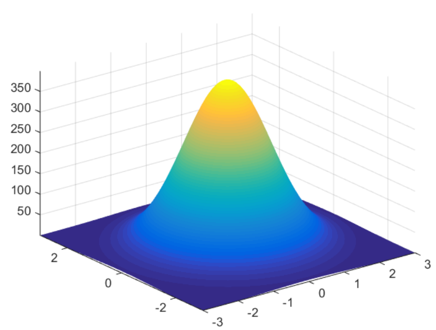
[DMNS06]

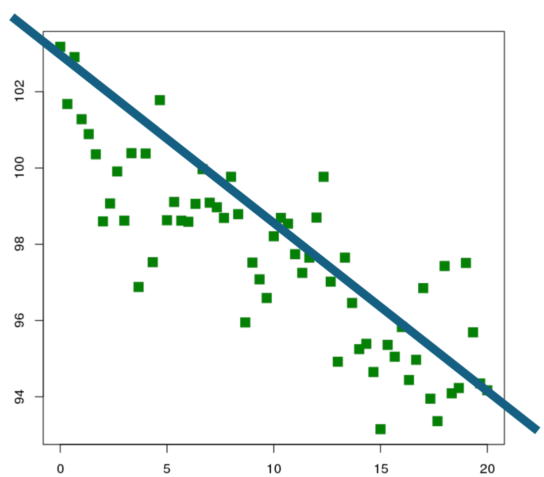
Privacy-Accuracy Tradeoff



Lots of recent progress!

- Estimate mean of bounded covariance distribution
- Learn a Gaussian
- Linear regression with/without condition-number dependence
- Learn a mixture of Gaussians
- Stochastic block model
- Graphon estimation
- …





What changed?

1. Different perspective: worst-case privacy, **average-case accuracy**.

   Example: privately release average of $X_1, \dots, X_n$ vs privately estimate the mean

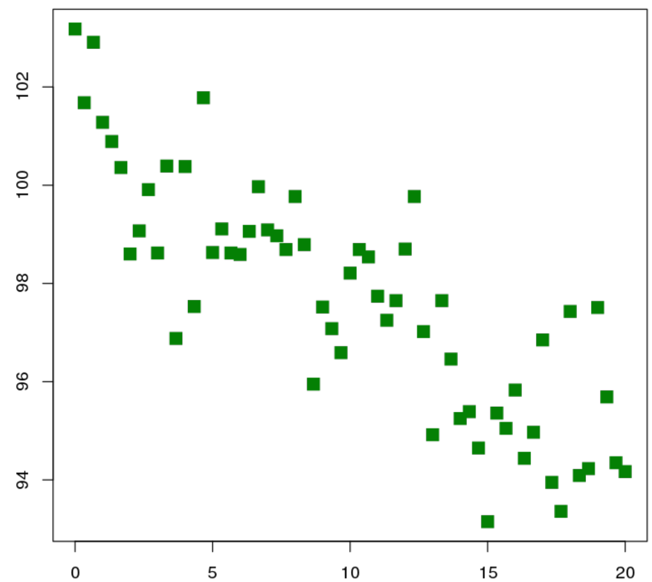
2. Renaissance in algorithmic robust statistics

Example 1: Mean Estimation, Bounded Covariance

Mean Estimation (Bdd Covariance)

Given: n iid samples $X_1, \dots, X_n$ from d-dimensional distribution D with $Cov(D) \preccurlyeq I$

Goal: Find $\hat{\mu} \in \mathbb{R}^d$ such that $\|\hat{\mu} - \mu(D)\| \leq \alpha$



Mean Estimation (Bdd Covariance)

Given: n iid samples $X_1, \dots, X_n$ from d-dimensional distribution D with $Cov(D) \preccurlyeq I$

Goal: Find $\hat{\mu} \in \mathbb{R}^d$ such that $\|\hat{\mu} - \mu(D)\| \leq \alpha$

"Differ on one individual": replace X_i with X_i' for a single $i \in [n]$

Mean Estimation (Bdd Covariance)

Given: n iid samples $X_1, \ldots, X_n$ from d-dimensional distribution D with $Cov(D) \preccurlyeq I$

Goal: Find $\hat{\mu} \in \mathbb{R}^d$ such that $\|\hat{\mu} - \mu(D)\| \leq \alpha$

Empirical mean: $n \asymp \frac{d}{\alpha^2}$, not private

Optimal tradeoff: $n \asymp \frac{d}{\varepsilon \alpha^2}$

[KSU20]

Mean Estimation (Bdd Covariance)

Given: n iid samples $X_1, \ldots, X_n$ from d-dimensional distribution D with $Cov(D) \preccurlyeq I$

Goal: Find $\hat{\mu} \in \mathbb{R}^d$ such that $\|\hat{\mu} - \mu(D)\| \leq \alpha$

Technical aside: for pure DP

Need assumption: $\|\mu\| \leq R$, known in advance

Naïve algos ("just add noise"): $n \gg \mathrm{poly}(R, d, 1/\varepsilon)$

Smarter algos: $n \gg \frac{d \log(R)}{\varepsilon} + \frac{d}{\varepsilon\alpha^2}$

[KV18, KLSU19]

Mean Estimation (Bdd Covariance)

Given: n iid samples $X_1, \ldots, X_n$ from d-dimensional distribution D with $Cov(D) \preccurlyeq I$

Goal: Find $\hat{\mu} \in \mathbb{R}^d$ such that $\|\hat{\mu} - \mu(D)\| \leq \alpha$

Technical aside: for approx. DP

~~Need assumption: $\|\mu\| \leq R$, known in advance~~

Instead: $n \gg \frac{\log 1/\delta}{\varepsilon}$

[KV18, KLSU19]

Given: n iid samples $X_1, \ldots, X_n$ from d-dimensional distribution D with $Cov(D) \preccurlyeq I$

Goal: Find $\hat{\mu} \in \mathbb{R}^d$ such that $\|\hat{\mu} - \mu(D)\| \leq \alpha$

Estimator	Samples*	Priv.?	Poly-Time?	Reference
Empirical mean	d/α^2	none		Folklore
Tournament	$d/\varepsilon\alpha^2$	pure		[KSU20]
Smart clip+noise	$d^{1.5}/\varepsilon\alpha^2$	pure		[KLSU19, KSU20]
Smart clip+noise	$\frac{d\sqrt{\log\frac{1}{\delta}}}{\varepsilon\alpha^2}$	appx.		[KLSU19, KSU20]
SoS Exp. Mech.	$d/\varepsilon\alpha^2$	pure		[**H**KM22]

Relies on poly-time robust mean estimator

*ignoring $\log d$, $\log 1/\alpha$ factors, $\log R$-dependence

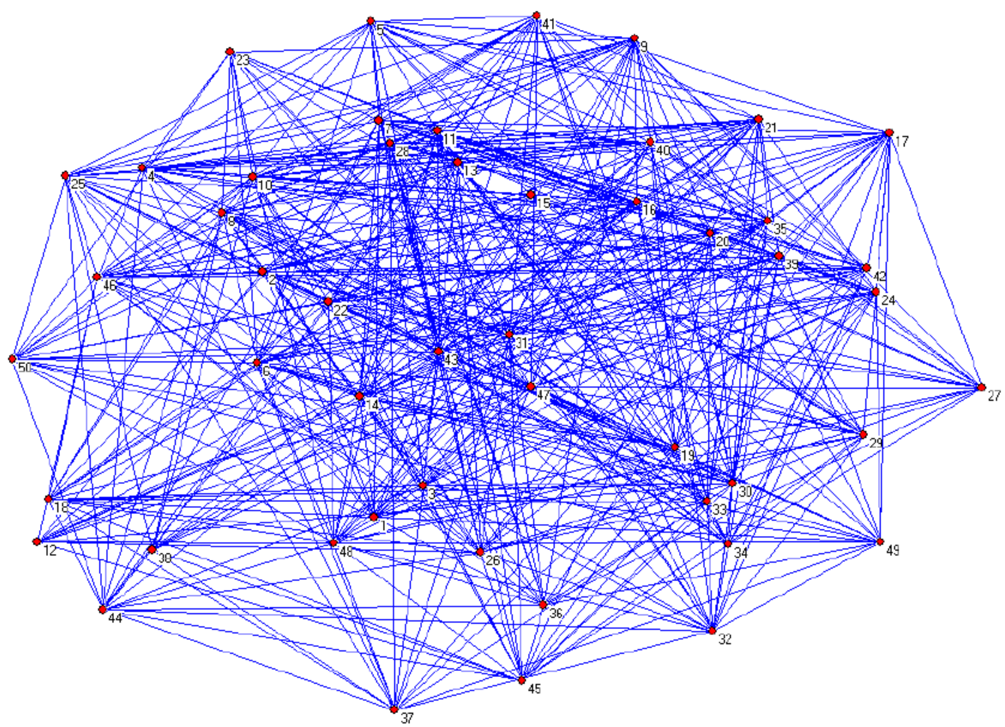
Example 2: Node-Private Graph Parameter Estimation

Graph Density Estimation

Given: Sample $G \sim G(n, p)$

Goal: Find $\hat{p}$ such that $|\hat{p} - p| \leq \alpha$

"Node privacy": G, G' differ on one *vertex*



Given: Sample $G \sim G(n, p)$

Goal: Find $\hat{p}$ such that $|\hat{p} - p| \leq \alpha$

"Node privacy": $\boldsymbol{G, G'}$ **differ on one *vertex***

Estimator	α^*	Priv.?	Poly-Time?	Reference
Edge count	$\frac{\sqrt{p}}{n}$			Folklore
Lipschitz Ext.	$\frac{\sqrt{p}}{n} + \frac{\sqrt{p}}{\varepsilon n^{1.5}}$			[BCSZ19]
Laplace noise	$\frac{\sqrt{p}}{\varepsilon n}$			(folklore)
Smooth sensitivity	$\frac{\sqrt{p}}{n} + \frac{\sqrt{p}}{\varepsilon n^{1.5}} + \frac{1}{\varepsilon^2 n^2}$			[SU19]
SoS Exp. Mech.	$\frac{\sqrt{p}}{n} + \frac{\sqrt{p}}{\varepsilon n^{1.5}}$			[CDHS24]

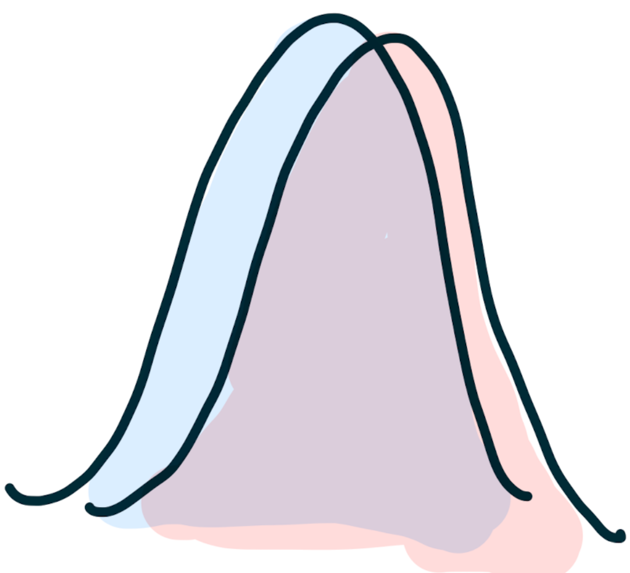
[AJKSZ22]

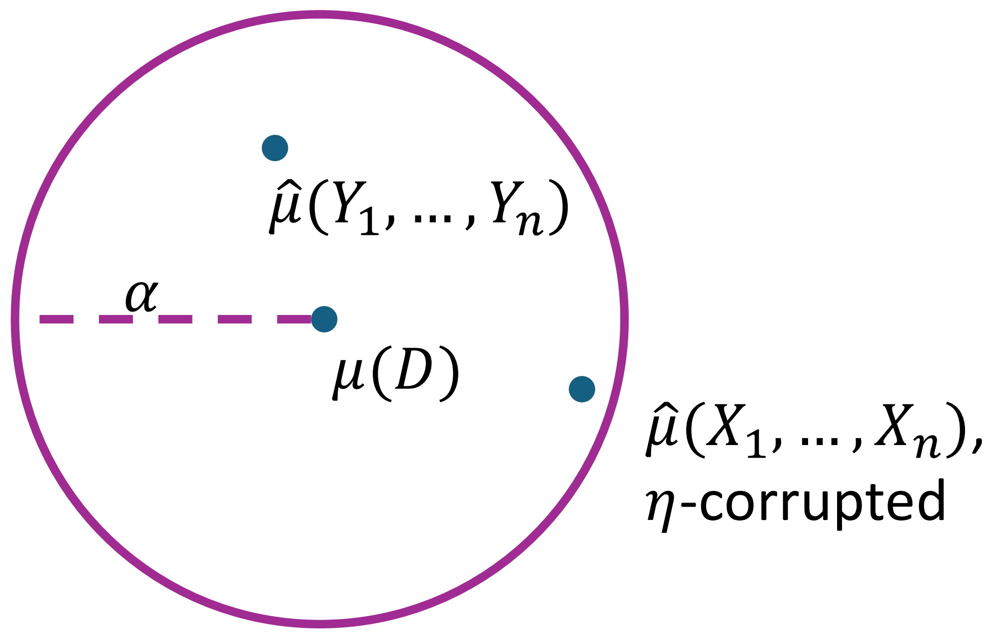
*ignoring logs

Robustness vs Privacy: Bird’s Eye View

Robustness vs Privacy: Intuitions

- Different measures of *stability when some inputs change*





Robustness vs Privacy: History

STOC 2009:

Differential Privacy and Robust Statistics

Cynthia Dwork
Microsoft Research

Jing Lei[*]
Department of Statistics

ABSTRACT

We show by means of several examples that robust statistical estimators present an excellent starting point for differentially private estimators. Our algorithms use a new paradigm for differentially private mechanisms, which we

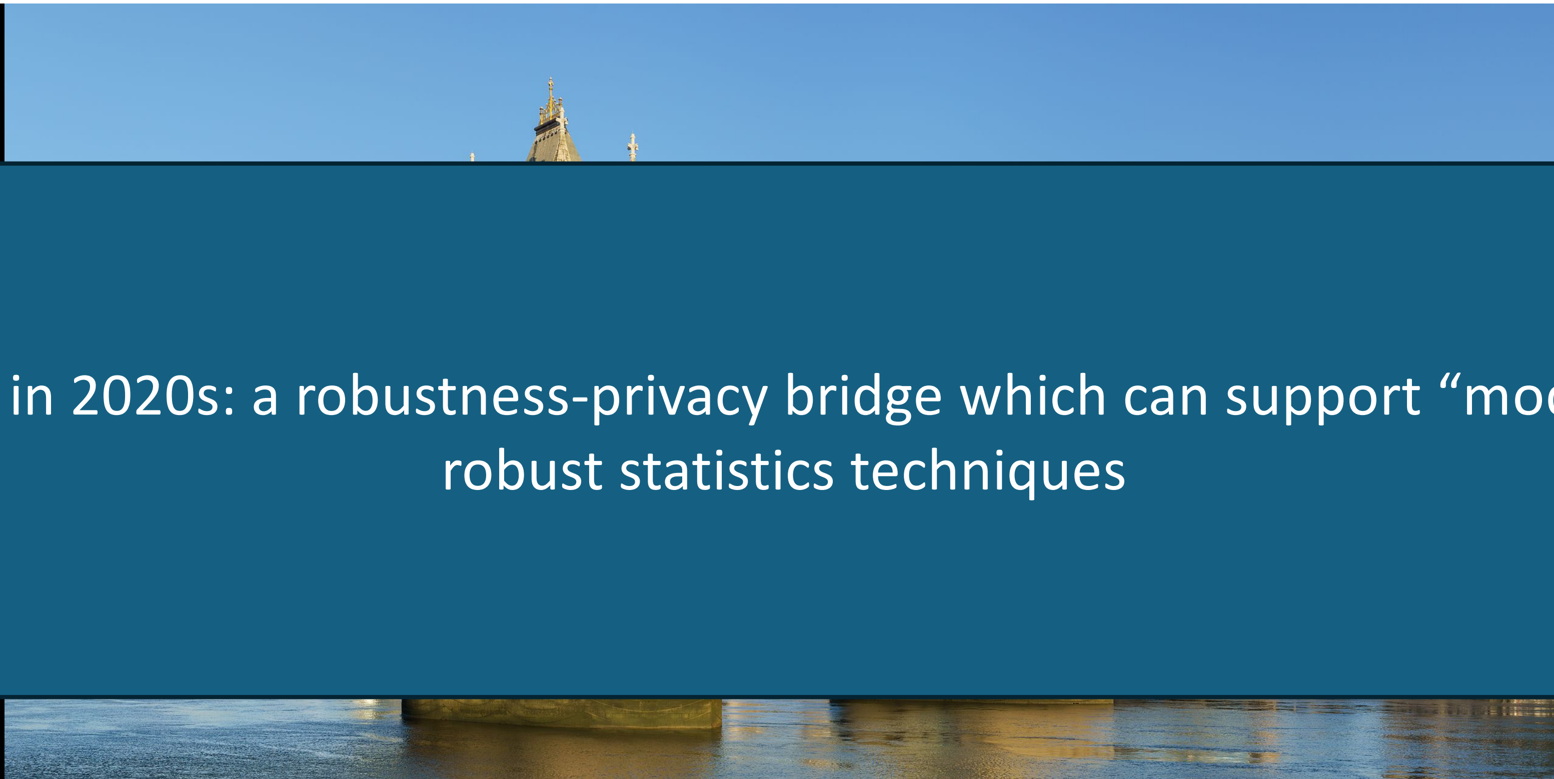
[S11, AD20, AMSSV20, LKO22, SV22, KMV22,…]

Yet, as of 2021, knew (nearly) optimal **robustness**-accuracy tradeoffs, in poly time, for

- Mean estimation
- Sparse mean estimation
- Learning Gaussian
- Linear regression
- Graph density estimation
- (many others)

And NOT optimal **privacy**-accuracy tradeoffs!

[AJKSZ22]

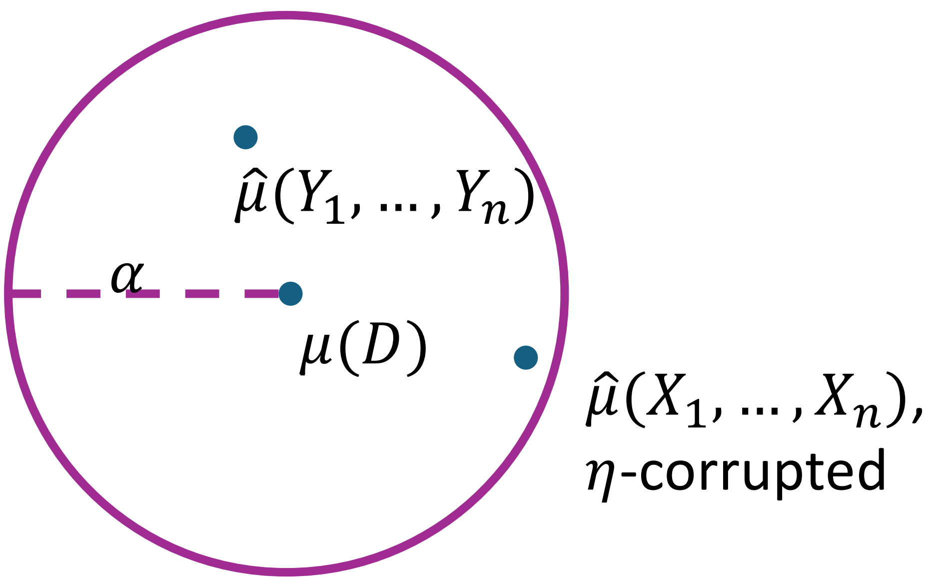


New in 2020s: a robustness-privacy bridge which can support “modern” robust statistics techniques

Robustness to Privacy

Two classes of techniques to leverage robust estimators

- Stability + noise
  - Typically not pure DP (good and bad…stay tuned)
  - [LWKO21, KMV22, CCEIST23, B**H**S23, B**H**HKLOPS24 LJWO24,…]



+

Robustness to Privacy

Two classes of techniques to leverage robust estimators

- Stability + noise (Gavin's talk)
  - Typically not pure DP (good and bad…stay tuned)
  - [LWKO21, KMV22, CCEIST23, B**H**S23, B**H**HKLOPS24 LJWO24,…]
  - Can produce (pretty) fast algorithms

Important difference from robustness: stable function $f: datasets \rightarrow outputs$ must satisfy bound on $||f(X) - f(X')||$ for **all** neighboring X, X'

Robustness to Privacy

Two classes of techniques to leverage robust estimators

- Stability + noise (Gavin's talk)
  - Typically not pure DP (good and bad…stay tuned)
  - [LWKO21, KMV22, CCEIST23, B**H**S23, B**H**HKLOPS24 LJWO24,…]
  - Can produce (pretty) fast algorithms
- Exponential mechanism, inverse sensitivity (Lydia and Mahbod's talks) [HT10, AD22, AUZ23, **H**KM22, **H**KMN23]
  - Algorithms often both private and robust
  - So far, only very slow (poly time) algorithms

Privacy to Robustness

Some private algorithms are robust **merely by virtue of their (very) strong privacy guarantees**.

(Many are not.)

Reasons to care:

- Avenue for robust algorithms (questionable…)
- Lens on how techniques should translate
- Transfer (computational) lower bounds
- Don't worry about robust **and** private algorithm design

[folklore, G**H**22]

Group Privacy and Robustness

Simple observation on group privacy

Suppose $M : datasets \rightarrow outputs$ satisfies (ε, δ)-DP and for an input X has:

$$\Pr_{\text{internal coins of } M}(M(X) \text{ is "good"}) \geq 1 - \beta$$

Then for every $X' \sim_{\eta n} X$,

$$\Pr_{\text{internal coins of } M}(M(X) \text{ is "good"}) \geq 1 - e^{\varepsilon \eta n}(\beta + \eta n \delta)$$

Simple observation on group privacy

Suppose $M : datasets \rightarrow outputs$ satisfies (ε, δ)-DP and for an input X has:

$$\Pr_{\text{internal coins of } M}(M(X) \text{ is ``good''}) \geq 1 - \beta$$

Then for every $X' \sim_{\eta n} X$,

$$\Pr_{\text{internal coins of } M}(M(X) \text{ is ``good''}) \geq 1 - e^{\varepsilon \eta n}(\beta + \eta n \delta)$$

So, can take η as large as $\min\left(\frac{\log \frac{1}{\beta}}{\varepsilon n}, \frac{\log \frac{1}{\delta}}{\varepsilon n}\right)$

So, can take η as large as $\min\left(\frac{\log\frac{1}{\beta}}{\varepsilon n}, \frac{\log\frac{1}{\delta}}{\varepsilon n}\right)$

So, can take η as large as $\min\left(\frac{\log\frac{1}{\beta}}{\varepsilon n}, \frac{\log\frac{1}{\delta}}{\varepsilon n}\right)$

What does it say for mean estimation?

Private mean estimation sample complexity (optimal):

$$n = \frac{d + \log\frac{1}{\beta}}{\varepsilon\alpha^2} + \frac{\log\frac{1}{\delta}}{\varepsilon}$$

$\rightarrow$ can take $\log 1/\beta$ as large as $\varepsilon\alpha^2 n$
$\rightarrow$ η as large as α^2
$\rightarrow$ $\alpha = \sqrt{\eta}$

So, can take η as large as $\min\left(\frac{\log\frac{1}{\beta}}{\varepsilon n}, \frac{\log\frac{1}{\delta}}{\varepsilon n}\right)$

What does it say for mean estimation?

Private mean estimation sample complexity (optimal):

$$n = \frac{d + \log\frac{1}{\beta}}{\varepsilon\alpha^2} + \frac{\log\frac{1}{\delta}}{\varepsilon}$$

$\rightarrow$ can take $\log 1/\beta$ as large as $\varepsilon\alpha^2 n$

$\rightarrow$ η as large as α^2

$\rightarrow$ $\alpha = \sqrt{\eta}$

Any sample-optimal private algorithm is robust!

So, can take η as large as $\min\left(\frac{\log\frac{1}{\beta}}{\varepsilon n}, \frac{\log\frac{1}{\delta}}{\varepsilon n}\right)$

What does it say for **Gaussian** mean estimation?

Private Gaussian mean estimation sample complexity (optimal):

$$n = \frac{d}{\alpha^2} + \frac{d + \log\frac{1}{\beta}}{\varepsilon\alpha} + \frac{\log\frac{1}{\delta}}{\varepsilon}$$

→ can take $\log 1/\beta$ as large as $\varepsilon\alpha n$

→ η as large as α

So, can take η as large as $\min\left(\frac{\log\frac{1}{\beta}}{\varepsilon n}, \frac{\log\frac{1}{\delta}}{\varepsilon n}\right)$

What does it say for **Gaussian** mean estimation?

Private Gaussian mean estimation sample complexity (optimal):

$$n = \frac{d}{\alpha^2} + \frac{d + \log\frac{1}{\beta}}{\varepsilon\alpha} + \frac{\log\frac{1}{\delta}}{\varepsilon}$$

→ can take $\log 1/\beta$ as large as $\varepsilon\alpha n$

→ η as large as α

Implies info-comp gap for private mean estimation [DKS17]

So, can take η as large as $\min\left(\frac{\log\frac{1}{\beta}}{\varepsilon n}, \frac{\log\frac{1}{\delta}}{\varepsilon n}\right)$

What does it say for clip+noise private mean estimation?

Old(er), approx.-DP mean estimator (clip+noise) (informal)

$$n \geq \frac{d \log 1/\beta}{\varepsilon \alpha^2}$$

$\rightarrow$ rearranges to $\frac{\log\frac{1}{\beta}}{\varepsilon n} \leq \frac{\alpha^2}{d}$

The Curious Tale of Covariance-Aware Mean Estimation

Covariance-Aware Mean Estimation (Gaussian case)

Samples $X_1, \ldots, X_n \sim N(\mu, \Sigma)$.

Goal: find $\hat{\mu}$ s.t. $\left\| \Sigma^{-\frac{1}{2}}(\hat{\mu} - \mu) \right\| \leq \alpha$

Empirical mean satisfies with d/α^2 samples

Private/robust?

Covariance-Aware Mean Estimation (Gaussian case)

Samples $X_1, \ldots, X_n \sim N(\mu, \Sigma)$.

Goal: find $\hat{\mu}$ s.t. $\left\| \Sigma^{-\frac{1}{2}} (\hat{\mu} - \mu) \right\| \leq \alpha$

Empirical mean satisfies with d/α^2 samples

estimate covariance (robustly/privately), then affine transform:

$\rightarrow$ $n \geq d^2$ samples to do robustly + poly time,

$\rightarrow$ $n \geq d^{1.5}$ samples to do privately

SQ lower bound: $n \geq \Omega(d^2)$ samples needed for robustness [D**H**PT]

Samples $X_1, \ldots, X_n \sim N(\mu, \Sigma)$.

Goal: find $\hat{\mu}$ s.t. $\left\| \Sigma^{-\frac{1}{2}} (\hat{\mu} - \mu) \right\| \leq \alpha$

[BGSUZ]: $n \geq \frac{d}{\alpha^2} + \frac{d}{\varepsilon \alpha} + \frac{\log \frac{1}{\delta}}{\varepsilon}$, exponential time

[B**H**S,DHK]: $n \geq \frac{d}{\alpha^2} + \frac{d \sqrt{\log \frac{1}{\delta}}}{\varepsilon \alpha} + \frac{d \log \frac{1}{\delta}}{\varepsilon}$, polynomial time

Samples $X_1, \ldots, X_n \sim N(\mu, \Sigma)$.

Goal: find $\hat{\mu}$ s.t. $\left\| \Sigma^{-\frac{1}{2}}(\hat{\mu} - \mu) \right\| \leq \alpha$

[BGSUZ]: $n \geq \frac{d}{\alpha^2} + \frac{d}{\varepsilon\alpha} + \frac{\log\frac{1}{\delta}}{\varepsilon}$, exponential time

[B**H**S,DHK]: $n \geq \frac{d}{\alpha^2} + \frac{d\sqrt{\log\frac{1}{\delta}}}{\varepsilon\alpha} + \frac{d\,\log\frac{1}{\delta}}{\varepsilon}$, polynomial time

Allows for samples-robustness tradeoff ηd^2

What's Next?

- DP is used in practice – are new algorithmic ideas helpful?
- Fast ("practical") algorithms with pure-DP guarantees
  - Generic technique to stabilize filters?
- Pure-DP algorithms for non-convex parameter spaces
  - Sparse mean estimation?