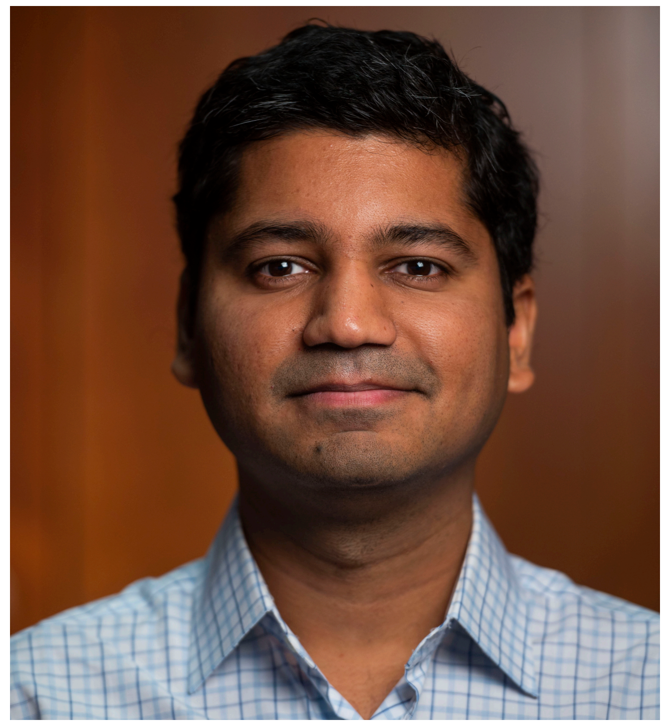
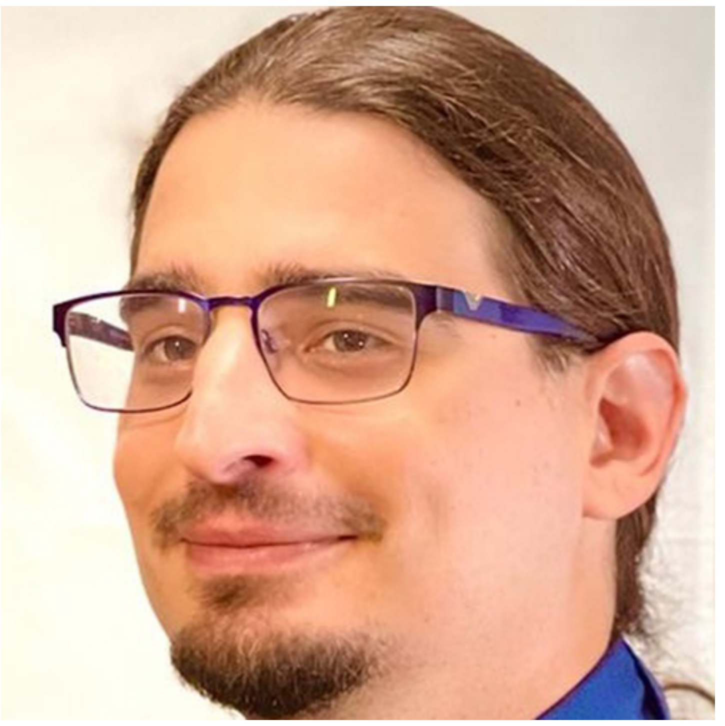
Robust Mean Estimation Against Oblivious Adversaries

Shuchen Li
CMU





Pravesh Kothari
CMU

Manolis Zampetakis
Yale

Robust statistics

Robust statistics

Huber's contamination model:

Robust statistics

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1 - \alpha)D + \alpha Z$

Robust statistics

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$

- D: "true" distribution (*inliers*)

Robust statistics

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$

- D: "true" distribution (*inliers*)
- Z: arbitrarily chosen by adversary (*outliers*)

Robust statistics

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1 - \alpha)D + \alpha Z$

- D: "true" distribution (*inliers*)
- Z: arbitrarily chosen by adversary (*outliers*)

Goal: estimate parameters of D (mean, covariance, regressor, …)

Robust mean estimation

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$

- $D = N(\mu, \Sigma)$
- Z: arbitrarily chosen by adversary

Goal: recover $\hat{\mu}$ with $||\mu - \hat{\mu}||_2 \leq \varepsilon$

Robust mean estimation

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$

- $D = N(\mu, \Sigma)$
- Z: arbitrarily chosen by adversary

Goal: recover $\hat{\mu}$ with $||\mu - \hat{\mu}||_2 \leq \varepsilon$

Fact: (information-theoretically) impossible if $\varepsilon < \left(\sqrt{\pi/2} - o(1)\right)\alpha$

[Diakonikolas-Kamath-Kane-Li-Moitra-Stewart'17]

Robust mean estimation

Huber's contamination model:

Input: $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$

- $D = N(\mu, \Sigma)$
- Z: arbitrarily chosen by adversary

Goal: recover $\hat{\mu}$ with $\|\mu - \hat{\mu}\|_2 \leq \varepsilon$

Fact: (information-theoretically) impossible if $\varepsilon < \left(\sqrt{\pi/2} - o(1)\right)\alpha$ 😭

[Diakonikolas-Kamath-Kane-Li-Moitra-Stewart'17]

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Our model (corruption before noise)

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Our model (corruption before noise)

1. $y'_1 = \cdots = y'_{(1-\alpha)n} = \mu,\ y'_{(1-\alpha)n+1}, \ldots, y'_{(1-\alpha)n+\alpha n}$ added by adversary

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Our model (corruption before noise)

1. $y'_1 = \cdots = y'_{(1-\alpha)n} = \mu,\ y'_{(1-\alpha)n+1}, \ldots, y'_{(1-\alpha)n+\alpha n}$ added by adversary

2. $y_i = y'_i + \eta_i$, the noise $\eta_i \overset{\text{i.i.d.}}{\sim} D(0)$

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Our model (corruption before noise)

1. $y'_1 = \cdots = y'_{(1-\alpha)n} = \mu, \; y'_{(1-\alpha)n+1}, \ldots, y'_{(1-\alpha)n+\alpha n}$ added by adversary
2. $y_i = y'_i + \eta_i$, the noise $\eta_i \overset{\text{i.i.d.}}{\sim} D(0)$

- Adversary is even *oblivious* of the *noise*

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Our model (corruption before noise)

1. $y'_1 = \cdots = y'_{(1-\alpha)n} = \mu,\ y'_{(1-\alpha)n+1}, \ldots, y'_{(1-\alpha)n+\alpha n}$ added by adversary
2. $y_i = y'_i + \eta_i$, the noise $\eta_i \overset{\text{i.i.d.}}{\sim} D(0)$

- Adversary is even *oblivious* of the *noise*

Our results: achieve ***arbitrarily*** high accuracy, given enough samples

Different models

Huber's contamination model:

- $y_1, y_2, \ldots, y_n \overset{\text{i.i.d.}}{\sim} (1-\alpha)D + \alpha Z$
- Adversary is *oblivious* of the inliers

Strong contamination model:

- $y_1, y_2, \ldots, y_{(1-\alpha)n} \overset{\text{i.i.d.}}{\sim} D$
- $y_{(1-\alpha)n+1}, \ldots, y_{(1-\alpha)n+\alpha n}$ added by adversary, *observing* the inliers

Our model (corruption before noise)

1. $y'_1 = \cdots = y'_{(1-\alpha)n} = \mu,\ y'_{(1-\alpha)n+1}, \ldots, y'_{(1-\alpha)n+\alpha n}$ added by adversary

2. $y_i = y'_i + \eta_i$, the noise $\eta_i \overset{\text{i.i.d.}}{\sim} D(0)$

- Adversary is even *oblivious* of the *noise*

Our results: achieve ***arbitrarily*** high accuracy, given enough samples

Why (algorithmic) robust statistics?

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression
   1) Choose n i.i.d. samples $x_i \sim N(0, I_d)$, and $y_i = \langle \ell^*, x_i \rangle + \eta_i$

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression
   1) Choose n i.i.d. samples $x_i \sim N(0, I_d)$, and $y_i = \langle \ell^*, x_i \rangle + \eta_i$
   2) Adversary observes samples, **replaces** α fraction of them

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression
   1) Choose n i.i.d. samples $x_i \sim N(0, I_d)$, and $y_i = \langle \ell^*, x_i \rangle + \eta_i$
   2) Adversary observes samples, **replaces** α fraction of them

↪ Algo for learning max of k linear models:

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression

   1) Choose n i.i.d. samples $x_i \sim N(0, I_d)$, and $y_i = \langle \ell^*, x_i \rangle + \eta_i$

   2) Adversary observes samples, **replaces** α fraction of them

↪ Algo for learning max of k linear models:

- $x_i \sim N(0, I_d)$, and $y_i = \max_{j \in [k]} \{ \langle \ell_j^*, x_i \rangle + \eta_j \}$

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression

   1) Choose n i.i.d. samples $x_i \sim N(0, I_d)$, and $y_i = \langle \ell^*, x_i \rangle + \eta_i$

   2) Adversary observes samples, **replaces** α fraction of them

↪ Algo for learning max of k linear models:

- $x_i \sim N(0, I_d)$, and $y_i = \max_{j \in [k]} \{ \langle \ell_j^*, x_i \rangle + \eta_j \}$
- Adversary replace with the max

Why (algorithmic) robust statistics?

1. **Classical problem** (e.g. [Huber'64], [Tukey'75])
2. **Applications:** e.g. robust linear regression

   1) Choose n i.i.d. samples $x_i \sim N(0, I_d)$, and $y_i = \langle \ell^*, x_i \rangle + \eta_i$

   2) Adversary observes samples, **replaces** α fraction of them

↪ Algo for learning max of k linear models:

- $x_i \sim N(0, I_d)$, and $y_i = \max_{j \in [k]} \{ \langle \ell_j^*, x_i \rangle + \eta_j \}$
- Adversary replace with the max
- Only $\sim 1/k$ fraction are uncorrupted, for $\ell^* = \ell_j^*$ for each j

Why oblivious adversaries?

Why oblivious adversaries?

1. Learning max of k linear models:
   - $x_i \sim N(0, I_d)$, and $y_i = \max_{j \in [k]} \{ \langle \ell_j^*, x_i \rangle + \eta_j \}$

Why oblivious adversaries?

1. Learning max of k linear models:
   - $x_i \sim N(0, I_d)$, and $y_i = \max_{j \in [k]} \{\langle \ell_j^*, x_i \rangle + \eta_j\}$
2. Max-affine regression:
   - $x_i \sim N(0, I_d)$, and $y_i = \max_{j \in [k]} \{\langle \ell_j^*, x_i \rangle\} + \eta_i$

Why oblivious adversaries?

1. Learning max of k linear models:
   - $x_i \sim N(0,I_d)$, and $y_i = \max_{j\in[k]}\{\langle \ell_j^*, x_i\rangle + \eta_j\}$
2. Max-affine regression:
   - $x_i \sim N(0,I_d)$, and $y_i = \max_{j\in[k]}\{\langle \ell_j^*, x_i\rangle\} + \eta_i$

Max *before* noise ↔ **corruption** *before* noise

Formulation

Formulation

1. Corrupted means:

Formulation

1. **Corrupted means:**

$$\{y_i'\} = \{\underbrace{\mu, \ldots, \mu}_{(1-\alpha)n}, z_1, \ldots, z_{\alpha n}\}$$

Formulation

1. **Corrupted means:**

$$\{y_i'\} = \{\underbrace{\mu, \ldots, \mu}_{(1-\alpha)n}, z_1, \ldots, z_{\alpha n}\}$$

2. **Add $N(0,I)$ noise:**

Formulation

1. **Corrupted means:**

$$\{y_i'\} = \{\underbrace{\mu, \ldots, \mu}_{(1-\alpha)n}, z_1, \ldots, z_{\alpha n}\}$$

2. **Add $N(0,I)$ noise:**

$$\{y_i\} \sim \{\underbrace{N(\mu,1), \ldots, N(\mu,1)}_{(1-\alpha)n}, N(z_1,1), \ldots, N(z_{\alpha n},1)\}$$

Formulation

1. **Corrupted means:**

$$\{y'_i\} = \{\underbrace{\mu, \ldots, \mu}_{(1-\alpha)n}, z_1, \ldots, z_{\alpha n}\}$$

2. **Add $N(0,I)$ noise:**

$$\{y_i\} \sim \{\underbrace{N(\mu,1), \ldots, N(\mu,1)}_{(1-\alpha)n}, N(z_1,1), \ldots, N(z_{\alpha n},1)\}$$

d-dim: project along each axis and run 1-dim algo for $\varepsilon/\sqrt{d}$ accuracy

Characteristic function

Characteristic function

Fact: For $X \sim N(\mu, \sigma^2)$, the *characteristic function* of X is

Characteristic function

Fact: For $X \sim N(\mu, \sigma^2)$, the *characteristic function* of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Characteristic function

Fact: For $X \sim N(\mu, \sigma^2)$, the *characteristic function* of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Averaging all the CFs of our input,

Characteristic function

Fact: For $X \sim N(\mu, \sigma^2)$, the *characteristic function* of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Averaging all the CFs of our input,

$$\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[e^{ity_j}] = (1-\alpha)e^{it\mu - \frac{1}{2}t^2} + \frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k - \frac{1}{2}t^2}$$

Characteristic function

Fact: For $X \sim N(\mu, \sigma^2)$, the *characteristic function* of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Averaging all the CFs of our input,

$$\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[e^{ity_j}]e^{\frac{1}{2}t^2} = (1-\alpha)e^{it\mu} + \frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}$$

Characteristic function

Fact: For $X \sim N(\mu, \sigma^2)$, the *characteristic function* of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

Averaging all the CFs of our input,

$$\frac{1}{n}\sum_{j=1}^{n} \mathbb{E}[e^{ity_j}]e^{\frac{1}{2}t^2} = (1-\alpha)e^{it\mu} + \frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}$$

Then apply Fourier transform to get frequency μ

Sparse Fourier Transform

Sparse Fourier Transform

SFT [Price-Song'16]: For *noisy 1-sparse* $x(t) = e^{itf} + g(t)$, $t \in [0, T]$, output f' that $|f - f'| \leq O(\mathcal{N}/T)$, if $\mathcal{N} = O(1)$, where

$$\mathcal{N} = \frac{1}{T}\int_0^T |g(t)|^2 \mathrm{d}t,$$

with $O(\log T)$ **samples** from $x(t)$ and in $O(\log(T)^2)$ **time**.

Sparse Fourier Transform

SFT [Price-Song'16]: For *noisy 1-sparse* $x(t) = e^{itf} + g(t)$, $t \in [0,T]$, output f' that $|f - f'| \leq O(\mathcal{N}/T)$, if $\mathcal{N} = O(1)$, where

$$\mathcal{N} = \frac{1}{T}\int_0^T |g(t)|^2 \mathrm{d}t,$$

with $O(\log T)$ **samples** from $x(t)$ and in $O(\log(T)^2)$ **time**.

Our Algo: Apply SFT on the empirical avg of CFs:

$$x(t) = \frac{1}{n}\sum_{j=1}^{n} e^{ity_j} e^{\frac{1}{2}t^2}$$

Analysis of the noise

Analysis of the noise

$$g(t) = \underbrace{\frac{1}{n}\sum_{j=1}^{n}(e^{ity_j} - \mathbb{E}[e^{ity_j}])e^{\frac{1}{2}t^2}}_{g_1(t)} + \underbrace{\frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}}_{g_2(t)}$$

Analysis of the noise

$$g(t) = \underbrace{\frac{1}{n}\sum_{j=1}^{n}(e^{ity_j} - \mathbb{E}[e^{ity_j}])e^{\frac{1}{2}t^2}}_{g_1(t)} + \underbrace{\frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}}_{g_2(t)}$$

1. Chernoff: $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot e^{T^2/2}\right)$

Analysis of the noise

$$g(t) = \underbrace{\frac{1}{n}\sum_{j=1}^{n}(e^{ity_j} - \mathbb{E}[e^{ity_j}])e^{\frac{1}{2}t^2}}_{g_1(t)} + \underbrace{\frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}}_{g_2(t)}$$

1. Chernoff: $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot e^{T^2/2}\right)$
2. $|g_2(t)| \leq \alpha$

Analysis of the noise

$$g(t) = \underbrace{\frac{1}{n}\sum_{j=1}^{n}(e^{ity_j} - \mathbb{E}[e^{ity_j}])e^{\frac{1}{2}t^2}}_{g_1(t)} + \underbrace{\frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}}_{g_2(t)}$$

1. Chernoff: $|g_1(t)| \le O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot e^{T^2/2}\right)$
2. $|g_2(t)| \le \alpha$

So we can only take $T = O(\sqrt{\log n})$, and $|\mu - \hat{\mu}| = O(1/\sqrt{\log n})$

Analysis of the noise

$$g(t) = \underbrace{\frac{1}{n}\sum_{j=1}^{n}(e^{ity_j} - \mathbb{E}[e^{ity_j}])e^{\frac{1}{2}t^2}}_{g_1(t)} + \underbrace{\frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}}_{g_2(t)}$$

1. Chernoff: $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot e^{T^2/2}\right)$
2. $|g_2(t)| \leq \alpha$

So we can only take $T = O(\sqrt{\log n})$, and $|\mu - \hat{\mu}| = O(1/\sqrt{\log n})$

- This gives $2^{O(\varepsilon^{-2})}$ sample/time complexity to achieve ε error

Analysis of the noise

$$g(t) = \underbrace{\frac{1}{n}\sum_{j=1}^{n}(e^{ity_j} - \mathbb{E}[e^{ity_j}])e^{\frac{1}{2}t^2}}_{g_1(t)} + \underbrace{\frac{1}{n}\sum_{k=1}^{\alpha n} e^{itz_k}}_{g_2(t)}$$

1. Chernoff: $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot e^{T^2/2}\right)$
2. $|g_2(t)| \leq \alpha$

So we can only take $T = O(\sqrt{\log n})$, and $|\mu - \hat{\mu}| = O(1/\sqrt{\log n})$

- This gives $2^{O(\varepsilon^{-2})}$ sample/time complexity to achieve ε error 😅

What if the noise is Laplace?

What if the noise is Laplace?

Fact: For $X \sim \text{Laplace}(\mu, b)$, the characteristic function of X is

What if the noise is Laplace?

Fact: For $X \sim \text{Laplace}(\mu, b)$, the characteristic function of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \frac{e^{it\mu}}{1 + b^2 t^2}$$

What if the noise is Laplace?

Fact: For $X \sim$ Laplace(μ, b), the characteristic function of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \frac{e^{it\mu}}{1 + b^2 t^2}$$

- $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot T^2\right)$

What if the noise is Laplace?

Fact: For $X \sim \text{Laplace}(\mu, b)$, the characteristic function of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \frac{e^{it\mu}}{1 + b^2 t^2}$$

- $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot T^2\right)$
- Now $T = O(n^{1/4-c})$

What if the noise is Laplace?

Fact: For $X \sim$ Laplace(μ, b), the characteristic function of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \frac{e^{it\mu}}{1 + b^2 t^2}$$

- $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot T^2\right)$
- Now $T = O(n^{1/4-c})$
- Sample/time complexity $O((1/\varepsilon)^{4+c})$

What if the noise is Laplace?

Fact: For $X \sim$ Laplace(μ, b), the characteristic function of X is

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \frac{e^{it\mu}}{1 + b^2 t^2}$$

- $|g_1(t)| \leq O\left(\frac{\sqrt{\log n}}{\sqrt{n}} \cdot T^2\right)$
- Now $T = O(n^{1/4-c})$
- Sample/time complexity $O((1/\varepsilon)^{4+c})$ 🧐

Back to Gaussian

Back to Gaussian

Fact: For $X \sim N(\mu, 1)$, the characteristic function of X^2 is

Back to Gaussian

Fact: For $X \sim N(\mu, 1)$, the characteristic function of X^2 is

$$\varphi_{X^2}(t) = \mathbb{E}\left[e^{itX^2}\right] = \frac{\exp(\frac{it\mu^2}{1-2it})}{(1-2it)^{1/2}}$$

Back to Gaussian

Fact: For $X \sim N(\mu, 1)$, the characteristic function of X^2 is

$$\varphi_{X^2}(t) = \mathbb{E}\left[e^{itX^2}\right] = \frac{\exp(\frac{it\mu^2}{1-2it})}{(1-2it)^{1/2}}$$

Substitute $\frac{t}{1-2it} \to t$,

Back to Gaussian

Fact: For $X \sim N(\mu, 1)$, the characteristic function of X^2 is

$$\varphi_{X^2}(t) = \mathbb{E}\left[e^{itX^2}\right] = \frac{\exp(\frac{it\mu^2}{1-2it})}{(1-2it)^{1/2}}$$

Substitute $\frac{t}{1-2it} \to t$,

$$\mathbb{E}\left[\exp\left(i\frac{t}{1+2it}X^2\right)\right](1+2it)^{-1/2} = e^{it\mu^2}$$

Back to Gaussian

Fact: For $X \sim N(\mu, 1)$, the characteristic function of X^2 is

$$\varphi_{X^2}(t) = \mathbb{E}\left[e^{itX^2}\right] = \frac{\exp(\frac{it\mu^2}{1-2it})}{(1-2it)^{1/2}}$$

Substitute $\frac{t}{1-2it} \to t$,

$$\mathbb{E}\left[\exp\left(i\frac{t}{1+2it}X^2\right)\right](1+2it)^{-1/2} = e^{it\mu^2}$$

norm $\sim \exp(X^2/2)$

too large to use concentration ineqs

Summary and open questions

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$
- **Laplace**: sample/time poly(d, ε^{-1})

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$
- **Laplace**: sample/time poly(d, ε^{-1})

Open questions:

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$
- **Laplace**: sample/time poly(d, ε^{-1})

Open questions:

- poly sample/time for Gaussian?

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$
- **Laplace**: sample/time poly(d, ε^{-1})

Open questions:

- poly sample/time for Gaussian?
- robust covariance estimation / linear regression in our setting?

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$
- **Laplace**: sample/time poly(d, ε^{-1})

Open questions:

- poly sample/time for Gaussian?
- robust covariance estimation / linear regression in our setting?
- list-decodable learning in our setting?

Summary and open questions

We give the first algorithm to achieve *arbitrary* accuracy in our *oblivious* model

- **Gaussian**: sample/time $2^{O(d/\varepsilon^2)}$
- **Laplace**: sample/time poly(d, ε^{-1})

Open questions:

- poly sample/time for Gaussian?
- robust covariance estimation / linear regression in our setting?
- list-decodable learning in our setting?

Thank you!