Gaps in Heavy-Tailed Statistics

Are There Information-Computation Gaps in High-Probability, Heavy-Tailed Estimation?

(unique to)

Sam Hopkins, UC Berkeley (→ MIT)

Based on joint work with Yeshwanth Cherapanamjeri, Tarun Kathuria, Prasad Raghavendra, and Nilesh Tripuraneni.

Measuring success probability in estimation (confidence intervals)

Given $X_1 \dots X_n \sim P_\theta$, find $\hat{\theta}$ s.t. $\|\theta - \hat{\theta}\| \leq \overbrace{r(n, d, \delta)}^{\text{"rate"}}$

with prob. $\geq 1-\delta$

no. samples

ambient dimension

failure rate

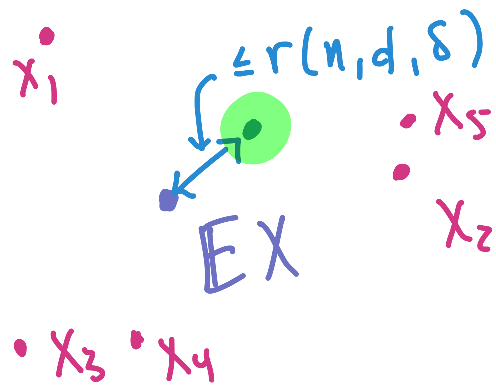
For $P_\theta \in \mathcal{P}$

Class of distributions

Our game: large class $\mathcal{P}$ (e.g. all distributions with $O(1)$ bdd. moments), but similar guarantees as if $\mathcal{P}$ contains Gaussians.

2 Illustrative Problems

- mean estimation $\|\cdot\| = \ell_2$ $\mathbb{E}X$
- covariance estimation $\|\cdot\| =$ operator norm $\mathbb{E}XX^T$



For this talk: all dist'ns have covariances $\Sigma \approx I$

$\mathrm{Tr}\,\Sigma = \Theta(d)$, $\|\Sigma\| = \Theta(1)$.

		mean	covar.
Gaussian	exponential time	$r^* = \sqrt{\frac{d}{n}} + \sqrt{\frac{\log 1/\delta}{n}}$	r^*
	polynomial time	r^*	r^*
O(1) bdd. moments	exponential time	r^* [LM '18]	r^* [MZ '19]
	$\mathrm{poly}(n,d,\frac{1}{\delta})$	r^*	r^* [CHKRT '20]
	$\mathrm{poly}(n,d,\log\frac{1}{\delta})$	r^* [H '18]	??

Open Problem: poly-time alg taking $X_1, \ldots, X_n \in \mathbb{R}^d$, output $\hat{\Sigma}$ s.t $\|\hat{\Sigma} - \Sigma\| \le O\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{\log 1/\delta}{n}}\right)$ w.p. $1-\delta$?

Or: rigorous evidence that no such alg exists?

- must apply only in small-δ regime ($\mathbb{E}\|\hat{\Sigma} - \Sigma\| \le \sqrt{\frac{d}{n}}$ possible)
- must apply only to non-Gaussian case (emp. covar. works for (sub)-Gaussian).

Rest of Talk:

1. Sketch of current state-of-the-art for covariance estimation

2. roadblocks to further improvement

(Nearly)-Optimal Covariance Estimation (Exp. Time)

$X_1 \dots X_n \in \mathbb{R}^d$, assume $\mathbb{E}X = 0$

1. $\underbrace{X_1, X_2,}\ \underbrace{\dots,}\ \underbrace{X_n}$

$\log\frac{1}{\delta}$ buckets

$Z_1 = \mathbb{E}_{i\sim B_1} X_i X_i^T \quad \dots \quad Z_{\log\frac{1}{\delta}} = \mathbb{E}_{i\sim B_{\log 1/\delta}} X_i X_i^T \in \mathbb{R}^{d\times d}$

[LM '19, MZ '19]

(Nearly)-Optimal Covariance Estimation (Exp. Time)

$X_1 \ldots X_n \in \mathbb{R}^d$, assume $\mathbb{E}X = 0$

1. $\underbrace{X_1, X_2,}\ \underbrace{\ldots,}\ \underbrace{X_n}$

$\log\frac{1}{\delta}$ buckets $\quad Z_1 = \mathbb{E}_{i \sim B_1} X_i X_i^T \quad \ldots \quad Z_{\log\frac{1}{\delta}} = \mathbb{E}_{i \sim B_{\log 1/\delta}} X_i X_i^T$

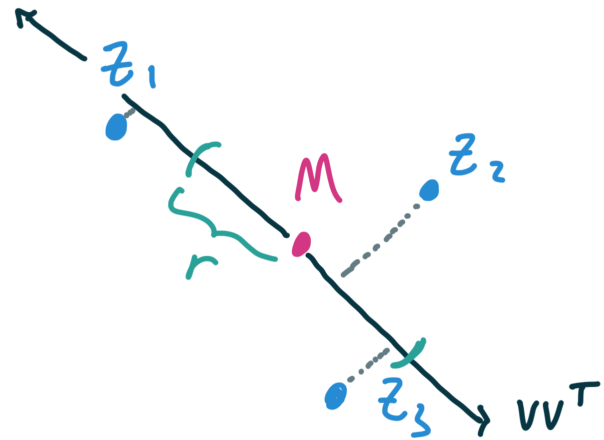
2. find M s.t. $\forall \|v\| = 1$,

$|v^T Z_i v - v^T M v| \le r$

for 60% of Z_i's,

for minimal r $(= r^*)$

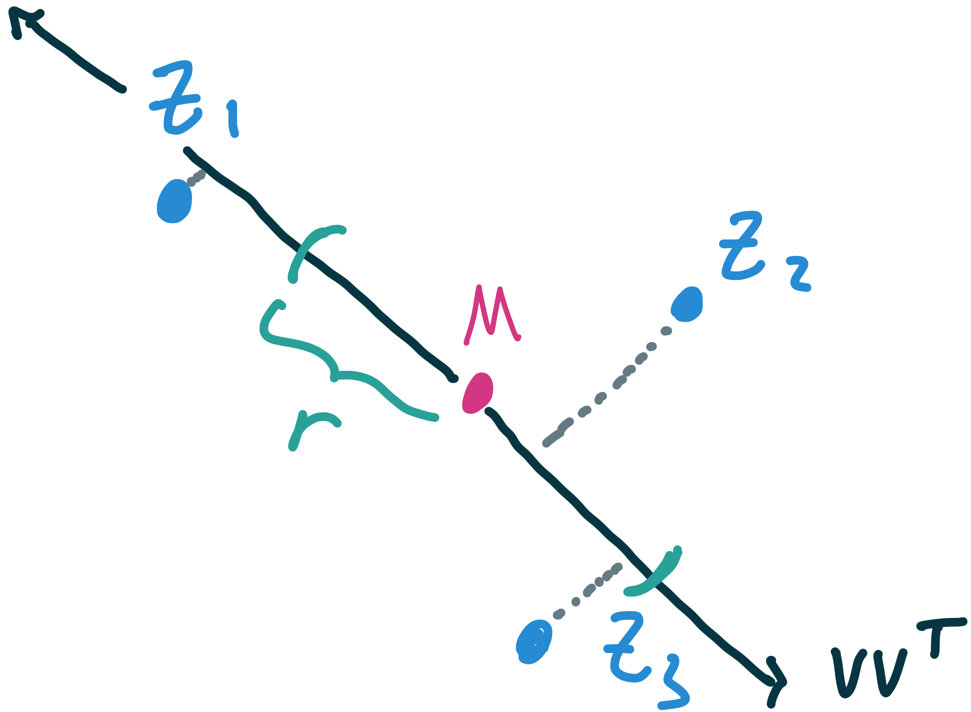
"Spectral r center"



[LM '19, MZ '19]

Theorem (LM, MZ): $\|M - \Sigma\| \leq \tilde{O}\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{\log 1/\delta}{n}}\right)$

w.p. $1-\delta$, assuming $\mathbb{E}\langle X, v\rangle^4 \leq O(\mathbb{E}\langle X, v\rangle^2)^2$



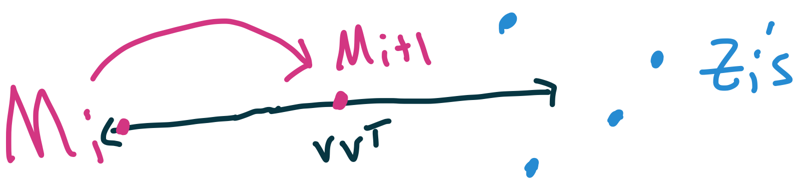
How to compute M? (spectral r center)

Theorem (CHKRT '20): In time $\mathrm{poly}(n, d, \log 1/\delta)$, can find M s.t. $\|M - \Sigma\| \le \tilde{O}\left(\frac{(\log \frac{1}{\delta})^{1/4} \cdot \sqrt{d}}{\sqrt{n}} + \sqrt{\frac{\log 1/\delta}{n}}\right)$ assuming SoS-certifiable 8^{th} moments.

$$\mathbb{E}\langle X, v\rangle^8 \le O\left(\mathbb{E}\langle X, v\rangle^2\right)^4 \qquad \sqrt{\frac{d}{n}} + \sqrt{\frac{\log 1/\delta}{n}}$$

Our Strategy: $M_1 \to M_2 \to \dots \to M$

- Attempt to <u>certify</u> that M_i is a spectral r center
- Success $\to$ output M_i
- Failure $\to$ witness v, update $M_{i+1} = M_i \pm vv^T$



[CFB'19, CHKRT'20]

- Attempt to certify that M_i is a spectral r center

Key lemma: for $r = \tilde{O}\left(\frac{(\log 1/\delta)^{1/4} \cdot \sqrt{d}}{\sqrt{n}} + \frac{\sqrt{\log 1/\delta}}{\sqrt{n}}\right)$, degree-8 SoS certifies, w.p. $1-\delta$, for $M = \Sigma$

But, true for $r = r^* = \sqrt{\frac{d}{n}} + \sqrt{\frac{\log 1/\delta}{n}}$

Can we certify for smaller r?

Conjectured hard dist'n: $(X_1, \dots, X_n)$ s.t.

- I is NOT a $\frac{(\log \frac{1}{\delta})^{0.24} \cdot \sqrt{d}}{\sqrt{n}} + \frac{\sqrt{\log 1/\delta}}{\sqrt{n}}$ center

- No $p(X_1 \dots X_n)$ of degree $(nd)^{o(1)}$ distinguishes from $X_1 \dots X_n \sim N(0, I)$

is a $\sqrt{\frac{d}{n}} + \sqrt{\frac{\log 1/\delta}{n}}$ center

Are there information-computation gaps unique to high-probability, heavy-tailed estimation?

$$\underbrace{X_1}_{N(0,\, I + vv^T)} , \underbrace{\ldots}\, \underbrace{\ldots}\, \underbrace{X_n}_{N(0,\, I - vv^T)}$$