Statistically Optimal Robust Mean and Covariance Estimation for Anisotropic Gaussians

Arshak Minasyan

CentraleSupélec - Université Paris-Saclay
Laboratoire Signaux et Systèmes (L2S)

Meeting in Mathematical Statistics 2023, Luminy, France

Decembre 18, 2023

arXiv preprint: A. Minasyan, N. Zhivotovskiy. Statistically Optimal Robust Mean and Covariance Estimation for Anisotropic Gaussians, *arXiv:2301.09024, 2023.*

Outline

Classical setting

We observe n vectors from $\mathbb{R}^d$ such that

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$$

with an unknown values of $\boldsymbol{\mu}^*, \boldsymbol{\Sigma}$.
The celebrated Borell, Tsirelson-Ibragimov-Sudakov Gaussian concentration states that w.p. $\geq 1 - \delta$, it holds

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i - \boldsymbol{\mu}^* \right\|_2 \leq \sqrt{\frac{\operatorname{Tr}(\boldsymbol{\Sigma})}{n}} + \sqrt{\frac{2\|\boldsymbol{\Sigma}\| \log(1/\delta)}{n}}.$$

Classical setting

We observe n vectors from $\mathbb{R}^d$ such that

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$$

with an unknown values of $\boldsymbol{\mu}^*, \boldsymbol{\Sigma}$.
The celebrated Borell, Tsirelson-Ibragimov-Sudakov Gaussian concentration states that w.p. $\geq 1 - \delta$, it holds

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i - \boldsymbol{\mu}^* \right\|_2 \leq \sqrt{\frac{\operatorname{Tr}(\boldsymbol{\Sigma})}{n}} + \sqrt{\frac{2\|\boldsymbol{\Sigma}\| \log(1/\delta)}{n}}.$$

Question: What happens if some "small" fraction of data is contaminated?

Classical setting

We observe n vectors from $\mathbb{R}^d$ such that

$$\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$$

with an unknown values of $\boldsymbol{\mu}^*, \boldsymbol{\Sigma}$.
The celebrated Borell, Tsirelson-Ibragimov-Sudakov Gaussian concentration states that w.p. $\geq 1 - \delta$, it holds

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i - \boldsymbol{\mu}^* \right\|_2 \leq \sqrt{\frac{\mathrm{Tr}(\boldsymbol{\Sigma})}{n}} + \sqrt{\frac{2\|\boldsymbol{\Sigma}\| \log(1/\delta)}{n}}.$$

Question: What happens if some "small" fraction of data is contaminated?
Goal: Find a robust estimator with the same rate as the sample mean and the *best possible* dependence on ε.

Contamination model

Definition (Gaussian Adversarial Contamination)

Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$ then we say that the distribution $\mathbf{P}_n$ of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ belongs to $\mathrm{GAC}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \varepsilon)$ with $\varepsilon \in (0, 1/2)$ satisfying

$$\big|\{i : \boldsymbol{X}_i \neq \boldsymbol{Y}_i\}\big| \leq \varepsilon n.$$

Outliers: $\mathcal{O} = \{i : \boldsymbol{X}_i \neq \boldsymbol{Y}_i\}$ *Inliers*: $\mathcal{I} = \{1, \ldots, n\} \setminus \mathcal{O}$.

Contamination model

Definition (Gaussian Adversarial Contamination)

Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$ then we say that the distribution $\mathbf{P}_n$ of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ belongs to $\mathrm{GAC}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \varepsilon)$ with $\varepsilon \in (0, 1/2)$ satisfying

$$\big|\{i : \boldsymbol{X}_i \neq \boldsymbol{Y}_i\}\big| \leq \varepsilon n.$$

Outliers: $\mathcal{O} = \{i : \boldsymbol{X}_i \neq \boldsymbol{Y}_i\}$ *Inliers*: $\mathcal{I} = \{1, \ldots, n\} \setminus \mathcal{O}$.

- GAC model allows both the set of **outliers** $\mathcal{O}$ and the **outliers themselves** to be random and to depend arbitrarily on the clean observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$.

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \text{med}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\text{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d}\sum_{i=1}^{n}\|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \mathrm{med}\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\mathrm{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

Robust !

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \mathrm{med}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\mathrm{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

Robust ! Optimal only in low dimensions, $d = O(1)$.

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \mathrm{med}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\mathrm{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

Robust ! Optimal only in low dimensions, $d = O(1)$.

- Tukey's median:

$$\widehat{\boldsymbol{\mu}}_n^{\mathrm{TM}} = \arg\max_{\boldsymbol{\mu}\in\mathbb{R}^d} \mathcal{D}_n(\boldsymbol{\mu}), \quad \mathcal{D}_n(\boldsymbol{\mu}) = \inf_{u\in\mathbb{S}^{d-1}} \sum_{i=1}^{n} \mathbb{1}(u^\top \boldsymbol{X}_i \le u^\top \boldsymbol{\mu})$$

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \mathrm{med}\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\mathrm{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

Robust ! Optimal only in low dimensions, $d = O(1)$.

- Tukey's median:

$$\widehat{\boldsymbol{\mu}}_n^{\mathrm{TM}} = \arg\max_{\boldsymbol{\mu}\in\mathbb{R}^d} \mathcal{D}_n(\boldsymbol{\mu}), \quad \mathcal{D}_n(\boldsymbol{\mu}) = \inf_{u\in\mathbb{S}^{d-1}} \sum_{i=1}^{n} \mathbb{1}(u^\top \boldsymbol{X}_i \le u^\top \boldsymbol{\mu})$$

Robust & optimal

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \mathrm{med}\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\mathrm{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

Robust ! Optimal only in low dimensions, $d = O(1)$.

- Tukey's median:

$$\widehat{\boldsymbol{\mu}}_n^{\mathrm{TM}} = \arg\max_{\boldsymbol{\mu}\in\mathbb{R}^d} \mathcal{D}_n(\boldsymbol{\mu}), \quad \mathcal{D}_n(\boldsymbol{\mu}) = \inf_{u\in\mathbb{S}^{d-1}} \sum_{i=1}^{n} \mathbb{1}(u^\top \boldsymbol{X}_i \le u^\top \boldsymbol{\mu})$$

Robust & optimal only when the <u>covariance is identical</u> and

~~Basic~~ estimators

- Sample mean:

$$\widehat{\boldsymbol{\mu}}_n = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i$$

Not robust ! One point can make the risk arbitrarily large.

- Coordinate-wise / geometric median:

$$\widehat{\boldsymbol{\mu}}_n = \text{med}\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\}, \qquad \widehat{\boldsymbol{\mu}}^{\text{GM}} \in \arg\min_{\boldsymbol{\mu}\in\mathbb{R}^d} \sum_{i=1}^{n} \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2.$$

Robust ! Optimal only in low dimensions, $d = O(1)$.

- Tukey's median:

$$\widehat{\boldsymbol{\mu}}_n^{\text{TM}} = \arg\max_{\boldsymbol{\mu}\in\mathbb{R}^d} \mathcal{D}_n(\boldsymbol{\mu}), \quad \mathcal{D}_n(\boldsymbol{\mu}) = \inf_{u\in\mathbb{S}^{d-1}} \sum_{i=1}^{n} \mathbb{1}(u^\top \boldsymbol{X}_i \leq u^\top \boldsymbol{\mu})$$

Robust & optimal only when the covariance is identical and takes exponential time to compute !

Lower bound

Using the results from [CGR18] and the lower bound in the outlier-free regime we have

$$\inf_{\widehat{\boldsymbol{\mu}}_n} \|\widehat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}^*\|_2 \geq c\|\boldsymbol{\Sigma}\|_{\text{op}}^{1/2}\left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \varepsilon\right)$$

holds with positive probability for some absolute constant $c > 0$, where

$$\mathbf{r}_{\boldsymbol{\Sigma}} \triangleq \frac{\operatorname{Tr}(\boldsymbol{\Sigma})}{\|\boldsymbol{\Sigma}\|_{\text{op}}}$$

is usually called *effective rank*.

Optimal robust mean estimation

Theorem (M., Zhivotovskiy 2023+)

Assume $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim \mathrm{GAC}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \varepsilon)$. *Let* $\varepsilon < c_1$, *then there is an estimator* $\widehat{\boldsymbol{\mu}}_n$ *satisfying, with probability at least* $1 - \delta$,

$$\|\widehat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}^*\|_2 \le c_2 \sqrt{\|\boldsymbol{\Sigma}\|} \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon \right),$$

where $c_1, c_2 > 0$ *are some absolute constants.*

Optimal robust mean estimation

Theorem (M., Zhivotovskiy 2023+)

Assume $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim \mathrm{GAC}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \varepsilon)$. *Let* $\varepsilon < c_1$, *then there is an estimator* $\widehat{\boldsymbol{\mu}}_n$ *satisfying, with probability at least* $1 - \delta$,

$$\|\widehat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}^*\|_2 \leq c_2 \sqrt{\|\boldsymbol{\Sigma}\|} \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon \right),$$

where $c_1, c_2 > 0$ *are some absolute constants.*

Let $\theta \sim \mathcal{N}(v, \beta^{-1} I_d)$, then

$$\widehat{\boldsymbol{\mu}}_n = \arg\min_{\nu \in \mathbb{R}^d} \sup_{v \in \mathbb{S}^{d-1}} \left| \mathbf{E}_{\rho_v} \operatorname{Med}(\langle \boldsymbol{X}_1, \theta \rangle, \ldots, \langle \boldsymbol{X}_n, \theta \rangle) - \langle \nu, v \rangle \right|.$$

tuning the parameter β

The parameter β is chosen by the Statistician in such a way that

$$\mathbf{r}_{\boldsymbol{\Sigma}}/10 \leq \beta \leq 10\mathbf{r}_{\boldsymbol{\Sigma}}.$$

tuning the parameter β

The parameter β is chosen by the Statistician in such a way that

$$\mathbf{r}_{\boldsymbol{\Sigma}}/10 \leq \beta \leq 10\mathbf{r}_{\boldsymbol{\Sigma}}.$$

To estimate β we need to estimate both $\|\boldsymbol{\Sigma}\|$ and $\mathrm{Tr}(\boldsymbol{\Sigma})$. [AZ22] provided an estimator ω such that $\|\boldsymbol{\Sigma}\|/4 \leq \omega \leq 4\|\boldsymbol{\Sigma}\|$. The estimation of $\mathrm{Tr}(\boldsymbol{\Sigma})$ reduces to mean estimation in $\mathbb{R}$ and using the estimator from [LM19] we have an estimator τ such that $\mathrm{Tr}(\boldsymbol{\Sigma})/2 \leq \tau \leq 2\mathrm{Tr}(\boldsymbol{\Sigma})$. Hence, this yields an estimator such that

$$\mathbf{r}_{\boldsymbol{\Sigma}}/8 \leq \frac{\tau}{\omega} \leq 8\mathbf{r}_{\boldsymbol{\Sigma}}.$$

An alternative approach for estimating β would be using Lepskii's method.

Coordinate-wise median of Gaussians

Let $X_1, \dots, X_n \sim \mathrm{GAC}(\mu^*, \sigma, \varepsilon)$, then, with probability at least $1 - \delta$,

$$|\mathrm{Med}(X_1, \dots, X_n) - \mu^*| \leq c\sigma \left(\sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon \right),$$

whenever $n \geq c \log(1/\delta)$ and ε is smaller than some absolute constant.

PAC-Bayesian Lemma

Lemma (Variational Inequality)

For a r.v. X on measurable space $\mathcal{X}$ and $\Theta \subset \mathbb{R}^d$. Let γ (prior) and ρ (posterior) be distributions on Θ, s.t. $\rho \ll \gamma$. Let $f : \mathcal{X} \times \Theta \to \mathbb{R}$ be such that $\mathbf{E}_X \exp(f(X, \theta)) < \infty$ γ-almost surely. Then, for all $\rho \ll \gamma$ simultaneously, we have, with a probability of at least $1 - e^{-t}$,

$$\mathbf{E}_\rho f(X, \theta) \le \mathbf{E}_\rho \log(\mathbf{E}_X \exp(f(X, \theta))) + \mathcal{KL}(\rho, \gamma) + t.$$

Gaussian quantiles

Lemma (Concentration for Gaussian quantiles)

Let $\varepsilon \in [0, 1/4]$*, and assume w.l.o.g. that* $(1/2 \pm \varepsilon)n$ *are integers. Let* $Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$*, then, for any* $t \geq 0$*,*

$$\Pr(|Y_{((1/2\pm\varepsilon)n)} - \Phi^{-1}(1/2 \pm \varepsilon)| \geq t) \leq 2\exp(-c_1 n t^2).$$

Equivalently,

$$\left\| Y_{((1/2\pm\varepsilon)n)} - \Phi^{-1}(1/2 \pm \varepsilon) \right\|_{\psi_2} \leq \frac{c_2}{\sqrt{n}}.$$

Half-normal quantiles

Lemma (Quantiles of the half-normal distribution)

Let $\varepsilon \in [0, 1/4]$. Assume w.l.o.g. that $(1/2 \pm \varepsilon)n$ are integers. Let $Y_1, \ldots, Y_n$ be a sample of independent half-normal random variables. Then,

$$\left\| Y_{((1/2\pm\varepsilon)n)} - \Phi_{\mathrm{H}}^{-1}(1/2 \pm \varepsilon) \right\|_{\psi_2} \leq \frac{c_1}{\sqrt{n}}.$$

χ_1^2 quantiles

Lemma (Quantiles of the χ_1^2 distribution)

Let $\varepsilon \in [0, 1/4]$. Assume w.l.o.g. that $(1/2 \pm \varepsilon)n$ are integers. Let $Y_1, \ldots, Y_n$ be a sample of independent χ_1^2 random variables. Then,

$$\left\| Y_{((1/2\pm\varepsilon)n)} - F_{\chi_1^2}^{-1}(1/2 \pm \varepsilon) \right\|_{\psi_1} \leq \frac{c_1}{\sqrt{n}}.$$

Classical setting. Covariance estimation

- In general, covariance estimation is a much more challenging problem.

Classical setting. Covariance estimation

- In general, covariance estimation is a much more challenging problem.
- We observe n vectors from $\mathbb{R}^d$ such that

$$\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \boldsymbol{\Sigma})$$

with an unknown value of $\boldsymbol{\Sigma}$.
The result of Koltchinskii & Lounici [KL17] states that w.p. $\geq 1 - \delta$, it holds

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\Sigma} \right\| \leq c_1 \|\boldsymbol{\Sigma}\| \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} \right),$$

whenever $n \geq c_2(\mathbf{r}_{\boldsymbol{\Sigma}} + \log(1/\delta))$.

Classical setting. Covariance estimation

- In general, covariance estimation is a much more challenging problem.
- We observe n vectors from $\mathbb{R}^d$ such that

$$\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \boldsymbol{\Sigma})$$

  with an unknown value of $\boldsymbol{\Sigma}$.
  The result of Koltchinskii & Lounici [KL17] states that w.p. $\geq 1 - \delta$, it holds

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\Sigma} \right\| \leq c_1 \|\boldsymbol{\Sigma}\| \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} \right),$$

  whenever $n \geq c_2(\mathbf{r}_{\boldsymbol{\Sigma}} + \log(1/\delta))$.

Question: What happens if some "small" fraction of data is contaminated?

Classical setting. Covariance estimation

- In general, covariance estimation is a much more challenging problem.
- We observe n vectors from $\mathbb{R}^d$ such that

$$\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(0, \boldsymbol{\Sigma})$$

with an unknown value of $\boldsymbol{\Sigma}$.
The result of Koltchinskii & Lounici [KL17] states that w.p. $\geq 1 - \delta$, it holds

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}_i^\top - \boldsymbol{\Sigma} \right\| \leq c_1 \|\boldsymbol{\Sigma}\| \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} \right),$$

whenever $n \geq c_2(\mathbf{r}_{\boldsymbol{\Sigma}} + \log(1/\delta))$.

Question: What happens if some "small" fraction of data is contaminated?
Goal: Find a robust estimator that has the same rate as sample covariance matrix and the dependence on ε is linear.

Robust covariance estimation

Theorem (M., Zhivotovskiy 2023+)

Assume that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GAC}(0, \boldsymbol{\Sigma}, \varepsilon)$. *Let* $\varepsilon < c_1$, *then there is an estimator* $\widehat{\boldsymbol{\Sigma}}$ *satisfying, with probability at least* $1 - \delta$,

$$\left\| \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\| \le c_2 \|\boldsymbol{\Sigma}\| \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon \right).$$

Robust covariance estimation

Theorem (M., Zhivotovskiy 2023+)

Assume that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} \mathrm{GAC}(0, \boldsymbol{\Sigma}, \varepsilon)$. *Let* $\varepsilon < c_1$, *then there is an estimator* $\widehat{\boldsymbol{\Sigma}}$ *satisfying, with probability at least* $1 - \delta$,

$$\left\| \widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma} \right\| \le c_2 \|\boldsymbol{\Sigma}\| \left(\sqrt{\frac{\mathbf{r}_{\boldsymbol{\Sigma}}}{n}} + \sqrt{\frac{\log(1/\delta)}{n}} + \varepsilon \right).$$

$$\widehat{\boldsymbol{\Sigma}} = \arg\min_{\Gamma \in \mathcal{H}} \sup_{v \in \mathbb{S}^{d-1}} \mathbf{E}_{\rho_v} \Big| \mathrm{Med}\left(|\langle \boldsymbol{X}_1, \theta \rangle|, \ldots, |\langle \boldsymbol{X}_n, \theta \rangle| \right) - \Phi^{-1}(3/4) \sqrt{\theta^\top \Gamma \theta} \Big|.$$

Concluding remarks

1. Gaussian assumption does not play a crucial role. Our results extend to distributions that are/have
   - Symmetry around the mean and spherical symmetry properties ($\langle X-\mu, v\rangle/\sqrt{v^\top \mathbf{\Sigma} v}$ for any $v \in S^{d-1}$ is independent of v).
   - For small enough ε and some $c > 0$
     $$|F^{-1}(1/2 \pm \varepsilon) - F^{-1}(1/2)| \leq c\varepsilon.$$
   - The density function f is separated from zero by an absolute constant for all $x \in [F^{-1}(1/2-\varepsilon), F^{-1}(1/2+\varepsilon)]$.
   - Sub-Gaussian tails.

Concluding remarks

1. Gaussian assumption does not play a crucial role. Our results extend to distributions that are/have
   - Symmetry around the mean and spherical symmetry properties ($\langle X - \mu, v\rangle / \sqrt{v^\top \mathbf{\Sigma} v}$ for any $v \in S^{d-1}$ is independent of v).
   - For small enough ε and some $c > 0$

     $$|F^{-1}(1/2 \pm \varepsilon) - F^{-1}(1/2)| \leq c\varepsilon.$$

   - The density function f is separated from zero by an absolute constant for all $x \in [F^{-1}(1/2 - \varepsilon), F^{-1}(1/2 + \varepsilon)]$.
   - Sub-Gaussian tails.
2. For mean estimation, it is a challenging problem to have a polynomial time algorithm with a linear dependence on ε, even when $\mathbf{\Sigma} = I_d$.

Concluding remarks

1. Gaussian assumption does not play a crucial role. Our results extend to distributions that are/have
   - Symmetry around the mean and spherical symmetry properties ($\langle X - \mu, v\rangle / \sqrt{v^\top \mathbf{\Sigma} v}$ for any $v \in S^{d-1}$ is independent of v).
   - For small enough ε and some $c > 0$

     $$|F^{-1}(1/2 \pm \varepsilon) - F^{-1}(1/2)| \leq c\varepsilon.$$

   - The density function f is separated from zero by an absolute constant for all $x \in [F^{-1}(1/2 - \varepsilon), F^{-1}(1/2 + \varepsilon)]$.
   - Sub-Gaussian tails.
2. For mean estimation, it is a challenging problem to have a polynomial time algorithm with a linear dependence on ε, even when $\mathbf{\Sigma} = I_d$.
3. For covariance estimation, there is no polynomial time algorithm achieving much weaker $\sqrt{\varepsilon}$-dependence.

Concluding remarks

1. Gaussian assumption does not play a crucial role. Our results extend to distributions that are/have
   - Symmetry around the mean and spherical symmetry properties ($\langle X - \mu, v\rangle / \sqrt{v^\top \boldsymbol{\Sigma} v}$ for any $v \in S^{d-1}$ is independent of v).
   - For small enough ε and some $c > 0$
   $$|F^{-1}(1/2 \pm \varepsilon) - F^{-1}(1/2)| \leq c\varepsilon.$$
   - The density function f is separated from zero by an absolute constant for all $x \in [F^{-1}(1/2 - \varepsilon), F^{-1}(1/2 + \varepsilon)]$.
   - Sub-Gaussian tails.
2. For mean estimation, it is a challenging problem to have a polynomial time algorithm with a linear dependence on ε, even when $\boldsymbol{\Sigma} = I_d$.
3. For covariance estimation, there is no polynomial time algorithm achieving much weaker $\sqrt{\varepsilon}$-dependence.
4. Lepskii's method as an alternative to estimate the unknown parameters.

Thank you !