Robust mean change point testing in high-dimensional data with heavy tails

Yudong Chen

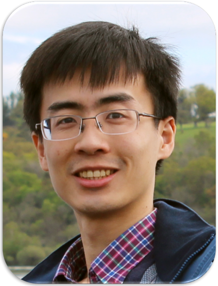
London School of Economics and Political Science

LSE Statistics Research Showcase 2023

6 June 2023

Collaborators

Mengchu Li



Tengyao Wang

Yi Yu

Reference

Li, M.*, Chen, Y.*, Wang, T. and Yu, Y. (2023) Robust mean change point testing in high-dimensional data with heavy tails. *arXiv preprint*, arXiv: 2305.18987.

Change point detection

Consider the model

$$\underset{\text{observations}}{\underset{\downarrow}{X}} = \underset{\text{signal}}{\underset{\downarrow}{\theta}} + \underset{\text{noise}}{\underset{\downarrow}{E}} \in \mathbb{R}^{p\times n}.$$

Entries of E are independent random variables with mean 0, variance 1 and distribution P_e.

Change point detection

Consider the model

$$
\underset{\text{observations}}{\underset{\downarrow}{X}} = \underset{\text{signal}}{\underset{\downarrow}{\theta}} + \underset{\text{noise}}{\underset{\downarrow}{E}} \in \mathbb{R}^{p \times n}.
$$

Entries of E are independent random variables with mean 0, variance 1 and distribution P_e.

Task: mean change point testing

$$
\begin{gathered}
H_0 : \text{no change in the columns of } \theta \\
\text{vs.} \\
H_1 : \exists \text{ a change in the columns of } \theta
\end{gathered}
$$

Change point detection – testing problem

Model: $X = \theta + E \in \mathbb{R}^{p \times n}$

- Null hypothesis H_0 (no change)

$$H_0 : \theta \in \Theta_0(p, n) := \left\{\theta : \theta_1 = \theta_2 = \ldots = \theta_n = \mu \in \mathbb{R}^p \text{ for some } \mu\right\}.$$

Change point detection – testing problem

Model: $X = \theta + E \in \mathbb{R}^{p \times n}$

- Null hypothesis H_0 (no change)

$$H_0 : \theta \in \Theta_0(p,n) := \big\{\theta : \theta_1 = \theta_2 = \ldots = \theta_n = \mu \in \mathbb{R}^p \text{ for some } \mu\big\}.$$

- Alternative hypothesis H_1 ($\exists$ change)

$$H_1 : \theta \in \Theta(p,n,s,\rho) := \bigcup_{t_0=1}^{n-1} \Theta^{(t_0)}(p,n,s,\rho)$$

where

$$\Theta^{(t_0)}(p,n,s,\rho) := \Big\{\theta : \theta_t = \mu_1 \text{ for } 1 \le t \le t_0,\ \theta_t = \mu_2 \text{ for } t_0 + 1 \le t \le n,\ \underbrace{\|\mu_1 - \mu_2\|_0}_{\text{sparsity level}} \le s,\ \underbrace{\frac{t_0(n-t_0)}{n}\|\mu_1 - \mu_2\|_2^2}_{\text{normalised signal strength}} \ge \rho^2\Big\}.$$

Minimax testing rate

Definition. Let Φ be the set of all test functions $\phi : \mathbb{R}^{p\times n} \to \{0,1\}$. Denote the minimax testing error

$$\begin{aligned}\mathcal{R}_{\mathcal{Q}}(\rho) &:= \inf_{\phi\in\Phi} \mathcal{R}_{\mathcal{Q}}(\rho,\phi) \\ &:= \inf_{\phi\in\Phi}\Big\{ \underbrace{\sup_{P_e\in\mathcal{Q}} \sup_{\theta\in\Theta_0(p,n)} \mathbb{E}\phi}_{\text{Type I error}} + \underbrace{\sup_{P_e\in\mathcal{Q}} \sup_{\theta\in\Theta(p,n,s,\rho)} \mathbb{E}(1-\phi)}_{\text{Type II error}} \Big\}.\end{aligned}$$

$v^*_{\mathcal{Q}}(p,n,s)$ is the **minimax testing rate** if

1. $\exists$ test ϕ, s.t. $\mathcal{R}_{\mathcal{Q}}(\rho,\phi) \le 1/2$ when $\rho^2 \gtrsim v^*_{\mathcal{Q}}(p,n,s)$,
2. $\forall$ test ϕ, have $\mathcal{R}_{\mathcal{Q}}(\rho,\phi) > 1/2$ when $\rho^2 \lesssim v^*_{\mathcal{Q}}(p,n,s)$.

Minimax testing rate

Definition. Let Φ be the set of all test functions $\phi : \mathbb{R}^{p\times n} \to \{0,1\}$. Denote the minimax testing error

$$\mathcal{R}_{\mathcal{Q}}(\rho) := \inf_{\phi\in\Phi} \mathcal{R}_{\mathcal{Q}}(\rho,\phi)$$

$$:= \inf_{\phi\in\Phi}\Big\{\underbrace{\sup_{P_e\in\mathcal{Q}}\sup_{\theta\in\Theta_0(p,n)}\mathbb{E}\phi}_{\text{Type I error}} + \underbrace{\sup_{P_e\in\mathcal{Q}}\sup_{\theta\in\Theta(p,n,s,\rho)}\mathbb{E}(1-\phi)}_{\text{Type II error}}\Big\}.$$

$v^*_{\mathcal{Q}}(p,n,s)$ is the **minimax testing rate** if

1. $\exists$ test ϕ, s.t. $\mathcal{R}_{\mathcal{Q}}(\rho,\phi) \le 1/2$ when $\rho^2 \gtrsim v^*_{\mathcal{Q}}(p,n,s)$,
2. $\forall$ test ϕ, have $\mathcal{R}_{\mathcal{Q}}(\rho,\phi) > 1/2$ when $\rho^2 \lesssim v^*_{\mathcal{Q}}(p,n,s)$.

- The distribution of each entry in E belongs to some class $\mathcal{Q}$.

Minimax testing rate

Definition. Let Φ be the set of all test functions $\phi : \mathbb{R}^{p\times n} \to \{0,1\}$. Denote the minimax testing error

$$\mathcal{R}_{\mathcal{Q}}(\rho) := \inf_{\phi\in\Phi} \mathcal{R}_{\mathcal{Q}}(\rho,\phi)$$

$$:= \inf_{\phi\in\Phi}\Big\{ \underbrace{\sup_{P_e\in\mathcal{Q}} \sup_{\theta\in\Theta_0(p,n)} \mathbb{E}\phi}_{\text{Type I error}} + \underbrace{\sup_{P_e\in\mathcal{Q}} \sup_{\theta\in\Theta(p,n,s,\rho)} \mathbb{E}(1-\phi)}_{\text{Type II error}} \Big\}.$$

$v^*_{\mathcal{Q}}(p,n,s)$ is the **minimax testing rate** if

1. $\exists$ test ϕ, s.t. $\mathcal{R}_{\mathcal{Q}}(\rho,\phi) \le 1/2$ when $\rho^2 \gtrsim v^*_{\mathcal{Q}}(p,n,s)$,
2. $\forall$ test ϕ, have $\mathcal{R}_{\mathcal{Q}}(\rho,\phi) > 1/2$ when $\rho^2 \lesssim v^*_{\mathcal{Q}}(p,n,s)$.

- The distribution of each entry in E belongs to some class $\mathcal{Q}$.
- Liu et al. (2021) derived the minimax testing rate for $\mathcal{Q} = \{N(0,1)\}$.
- Heavy-tailed distributions in $\mathcal{Q}$?

Heavy-tailed distributions

Two types of heavy-tailedness:

Definition ($\mathcal{G}_{\alpha,K}$ class). For any $P \in \mathcal{G}_{\alpha,K}$ and r.v. $W \sim P$,

$$\mathbb{E}W = 0, \quad \mathbb{E}W^2 = 1 \quad \text{and} \quad \mathbb{E}\exp\{|W/K|^\alpha\} \leq 2.$$

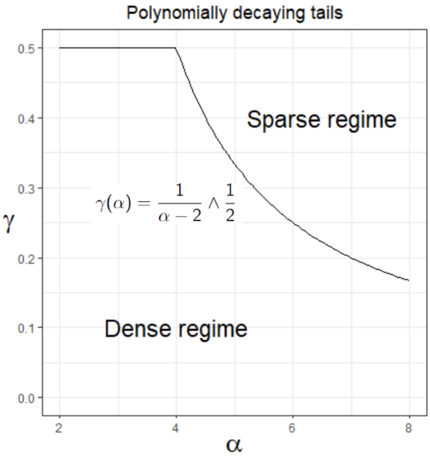
Sub-Weibull distributions of order α; possessing exponentially-decaying tails

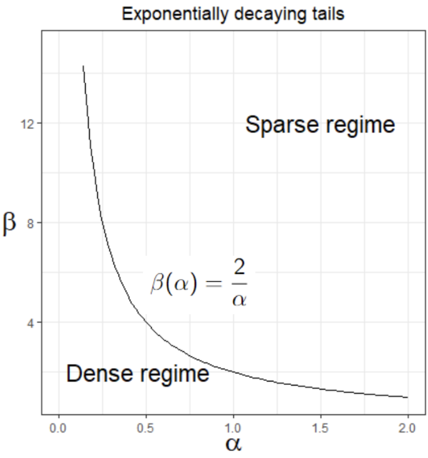
Definition ($\mathcal{P}_{\alpha,K}$ class). For any $P \in \mathcal{P}_{\alpha,K}$ and r.v. $W \sim P$,

$$\mathbb{E}W = 0, \quad \mathbb{E}W^2 = 1 \quad \text{and} \quad \mathbb{E}|W/K|^\alpha \leq 1.$$

Distributions with **finite α-th moment**; possessing polynomially-decaying tails

Main results – transition boundary





$$s^*_{\mathcal{P}} = p^{\frac{1}{2}} p^{-\gamma(\alpha)}$$

$$s^*_{\mathcal{G}} = p^{\frac{1}{2}} \log^{-\beta(\alpha)}(ep)$$

Main results – minimax rates

- Minimax rate upper bound v^{U}: construct a test procedure ϕ, such that $\mathcal{R}_{\mathcal{Q}}(\rho, \phi) \leq 1/2$ when $\rho^2 \geq v^{\mathrm{U}}$.
- Minimax rate lower bound v^{L}: usually via Le Cam's two point method.

		Upper bound	Lower bound
$\mathcal{G}^{\otimes}_{\alpha,K}$	Dense	(i) $\sqrt{p \log\log(8n)} + \log\log(8n)$	(ii) $\sqrt{p(\log\log(8n))^{\omega_1}} + \log\log(8n)$
	Sparse	(iii) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$	(iv) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$
$\mathcal{P}^{\otimes}_{\alpha,K}$	Dense	(v) $p^{(2/\alpha)\vee(1/2)} \log\log(8n)$	(vi) $p^{(2/\alpha)\vee(1/2)}\sqrt{(\log\log(8n))^{\omega_2}} + \log\log(8n)$
	Sparse	(vii) $s(p/s)^{2/\alpha} + \log\log(8n)$	(viii) $s(p/s)^{2/\alpha} + \log\log(8n)$

Note: $\omega_1 = \mathbb{1}_{\{s>\sqrt{p\log\log(8n)}\}}$ and $\omega_2 = \mathbb{1}_{\{s>\sqrt{p\log\log(8n)} \text{ and } \alpha\geq 4\}}$.

Main results – minimax rates

- Minimax rate upper bound v^{U}: construct a test procedure ϕ, such that $\mathcal{R}_{\mathcal{Q}}(\rho, \phi) \leq 1/2$ when $\rho^2 \geq v^{\mathrm{U}}$.
- Minimax rate lower bound v^{L}: usually via Le Cam's two point method.

		Upper bound	Lower bound
$\mathcal{G}^{\otimes}_{\alpha,K}$	Dense	(i) $\sqrt{p \log\log(8n)} + \log\log(8n)$	(ii) $\sqrt{p(\log\log(8n))^{\omega_1}} + \log\log(8n)$
	Sparse	(iii) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$	(iv) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$
$\mathcal{P}^{\otimes}_{\alpha,K}$	Dense	(v) $p^{(2/\alpha)\vee(1/2)} \log\log(8n)$	(vi) $p^{(2/\alpha)\vee(1/2)}\sqrt{(\log\log(8n))^{\omega_2}} + \log\log(8n)$
	Sparse	(vii) $s(p/s)^{2/\alpha} + \log\log(8n)$	(viii) $s(p/s)^{2/\alpha} + \log\log(8n)$

Note: $\omega_1 = \mathbb{1}_{\{s > \sqrt{p \log\log(8n)}\}}$ and $\omega_2 = \mathbb{1}_{\{s > \sqrt{p \log\log(8n)} \text{ and } \alpha \geq 4\}}$.

Matching rates!

Main results – minimax rates

- Minimax rate upper bound v^{U}: construct a test procedure ϕ, such that $\mathcal{R}_{\mathcal{Q}}(\rho, \phi) \leq 1/2$ when $\rho^2 \geq v^{\mathrm{U}}$.
- Minimax rate lower bound v^{L}: usually via Le Cam's two point method.

		Upper bound	Lower bound
$\mathcal{G}^{\otimes}_{\alpha,K}$	Dense	(i) $\sqrt{p \log\log(8n)} + \log\log(8n)$	(ii) $\sqrt{p(\log\log(8n))^{\omega_1}} + \log\log(8n)$
	Sparse	(iii) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$	(iv) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$
$\mathcal{P}^{\otimes}_{\alpha,K}$	Dense	(v) $p^{(2/\alpha)\vee(1/2)} \log\log(8n)$	(vi) $p^{(2/\alpha)\vee(1/2)} \sqrt{(\log\log(8n))^{\omega_2}} + \log\log(8n)$
	Sparse	(vii) $s(p/s)^{2/\alpha} + \log\log(8n)$	(viii) $s(p/s)^{2/\alpha} + \log\log(8n)$

Note: $\omega_1 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)}\right\}}$ and $\omega_2 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)} \text{ and } \alpha \geq 4\right\}}$.

A minimax gap of $\sqrt{\log\log n}$

Main results – minimax rates

- Minimax rate upper bound v^{U}: construct a test procedure ϕ, such that $\mathcal{R}_{\mathcal{Q}}(\rho, \phi) \leq 1/2$ when $\rho^2 \geq v^{\mathrm{U}}$.
- Minimax rate lower bound v^{L}: usually via Le Cam's two point method.

		Upper bound	Lower bound
$\mathcal{G}^{\otimes}_{\alpha,K}$	Dense	(i) $\sqrt{p \log\log(8n)} + \log\log(8n)$	(ii) $\sqrt{p(\log\log(8n))^{\omega_1}} + \log\log(8n)$
	Sparse	(iii) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$	(iv) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$
$\mathcal{P}^{\otimes}_{\alpha,K}$	Dense	(v) $p^{(2/\alpha)\vee(1/2)} \log\log(8n)$	(vi) $p^{(2/\alpha)\vee(1/2)}\sqrt{(\log\log(8n))^{\omega_2}} + \log\log(8n)$
	Sparse	(vii) $s(p/s)^{2/\alpha} + \log\log(8n)$	(viii) $s(p/s)^{2/\alpha} + \log\log(8n)$

Note: $\omega_1 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)}\right\}}$ and $\omega_2 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)} \text{ and } \alpha \geq 4\right\}}$.

A minimax gap of $\log\log n$

Main results – minimax rates

- Minimax rate upper bound v^{U}: construct a test procedure ϕ, such that $\mathcal{R}_{\mathcal{Q}}(\rho, \phi) \leq 1/2$ when $\rho^2 \geq v^{\mathrm{U}}$.
- Minimax rate lower bound v^{L}: usually via Le Cam's two point method.

		Upper bound	Lower bound
$\mathcal{G}^{\otimes}_{\alpha,K}$	Dense	(i) $\sqrt{p \log\log(8n)} + \log\log(8n)$	(ii) $\sqrt{p(\log\log(8n))^{\omega_1}} + \log\log(8n)$
	Sparse	(iii) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$	(iv) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$
$\mathcal{P}^{\otimes}_{\alpha,K}$	Dense	(v) $p^{(2/\alpha)\vee(1/2)} \log\log(8n)$	(vi) $p^{(2/\alpha)\vee(1/2)} \sqrt{(\log\log(8n))^{\omega_2}} + \log\log(8n)$
	Sparse	(vii) $s(p/s)^{2/\alpha} + \log\log(8n)$	(viii) $s(p/s)^{2/\alpha} + \log\log(8n)$

Note: $\omega_1 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)}\right\}}$ and $\omega_2 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)} \text{ and } \alpha \geq 4\right\}}$.

Sub-Weibull $\mathcal{G}_{\alpha,K}$ – dense

- Consider a dyadic grid $\mathcal{T} := \left\{1, 2, 4, \ldots, 2^{\lfloor \log_2(n/2) \rfloor}\right\}$ and CUSUM-type statistics
$$Y_t := \frac{\sum_{i=1}^{t} X_i - \sum_{i=1}^{t} X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$
- Aggregation across coordinates:
$$A_t := \sum_{j=1}^{p} \left(Y_t^2(j) - 1\right).$$
- Test:
$$\phi_{\mathcal{G},\text{dense}} := \mathbb{1}_{\{\max_{t \in \mathcal{T}} A_t > r\}}.$$

Sub-Weibull $\mathcal{G}_{\alpha,K}$ – dense

- Consider a dyadic grid $\mathcal{T} := \left\{1, 2, 4, \ldots, 2^{\lfloor \log_2(n/2) \rfloor}\right\}$ and CUSUM-type statistics
$$Y_t := \frac{\sum_{i=1}^t X_i - \sum_{i=1}^t X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$
- Aggregation across coordinates:
$$A_t := \sum_{j=1}^p \left(Y_t^2(j) - 1\right).$$
- Test:
$$\phi_{\mathcal{G},\mathrm{dense}} := \mathbb{1}_{\{\max_{t \in \mathcal{T}} A_t > r\}}.$$
- $\mathcal{R}_{\mathcal{G}}(\rho, \phi_{\mathcal{G},\mathrm{dense}}) \leq 1/2$ as long as $\rho^2 \gtrsim \sqrt{p \log\log(8n)} + \log\log(8n)$.

Sub-Weibull $\mathcal{G}_{\alpha,K}$ – sparse

- Recall that for $t \in \mathcal{T}$, we compute

$$Y_t := \frac{\sum_{i=1}^{t} X_i - \sum_{i=1}^{t} X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$

and

$$A_t := \sum_{j=1}^{p} \big(Y_t^2(j) - 1\big).$$

Sub-Weibull $\mathcal{G}_{\alpha,K}$ – sparse

- Recall that for $t \in \mathcal{T}$, we compute

$$Y_t := \frac{\sum_{i=1}^{t} X_i - \sum_{i=1}^{t} X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$

and

$$A_{t,a} := \sum_{j=1}^{p} \big(Y_t^2(j) - 1\big) \mathbb{1}_{\{|Y_t(j)| \geq a\}}.$$

Thresholding step

Sub-Weibull $\mathcal{G}_{\alpha,K}$ – sparse

- Recall that for $t \in \mathcal{T}$, we compute

$$Y_t := \frac{\sum_{i=1}^{t} X_i - \sum_{i=1}^{t} X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$

and

$$A_{t,a} := \sum_{j=1}^{p} \big(Y_{t,1}^2(j) - 1 \big) \mathbb{1}_{\{|Y_{t,2}(j)| \geq a\}}.$$

Thresholding step + sample splitting

Sub-Weibull $\mathcal{G}_{\alpha,K}$ – sparse

- Recall that for $t \in \mathcal{T}$, we compute

$$Y_t := \frac{\sum_{i=1}^t X_i - \sum_{i=1}^t X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$

and

$$A_{t,a} := \sum_{j=1}^p \left(Y_{t,1}^2(j) - 1\right) \mathbb{1}_{\{|Y_{t,2}(j)| \geq a\}}.$$

Thresholding step + sample splitting

- Test:

$$\phi_{\mathcal{G},\text{sparse}} := \mathbb{1}_{\{\max_{t\in\mathcal{T}} A_{t,a} > r\}}.$$

- $\mathcal{R}_{\mathcal{G}}(\rho, \phi_{\mathcal{G},\text{sparse}}) \leq 1/2$ as long as $\rho^2 \gtrsim s \log^{2/\alpha}(ep/s) + \log\log(8n)$.

Main results – minimax rates

- Minimax rate upper bound v^{U}: construct a test procedure ϕ, such that $\mathcal{R}_{\mathcal{Q}}(\rho, \phi) \leq 1/2$ when $\rho^2 \geq v^{\mathrm{U}}$.
- Minimax rate lower bound v^{L}: usually via Le Cam's two point method.

		Upper bound	Lower bound
$\mathcal{G}^{\otimes}_{\alpha,K}$	Dense	(i) $\sqrt{p \log\log(8n)} + \log\log(8n)$	(ii) $\sqrt{p(\log\log(8n))^{\omega_1}} + \log\log(8n)$
	Sparse	(iii) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$	(iv) $s \log^{2/\alpha}(ep/s) + \log\log(8n)$
$\mathcal{P}^{\otimes}_{\alpha,K}$	Dense	(v) $p^{(2/\alpha)\vee(1/2)} \log\log(8n)$	(vi) $p^{(2/\alpha)\vee(1/2)} \sqrt{(\log\log(8n))^{\omega_2}} + \log\log(8n)$
	Sparse	(vii) $s(p/s)^{2/\alpha} + \log\log(8n)$	(viii) $s(p/s)^{2/\alpha} + \log\log(8n)$

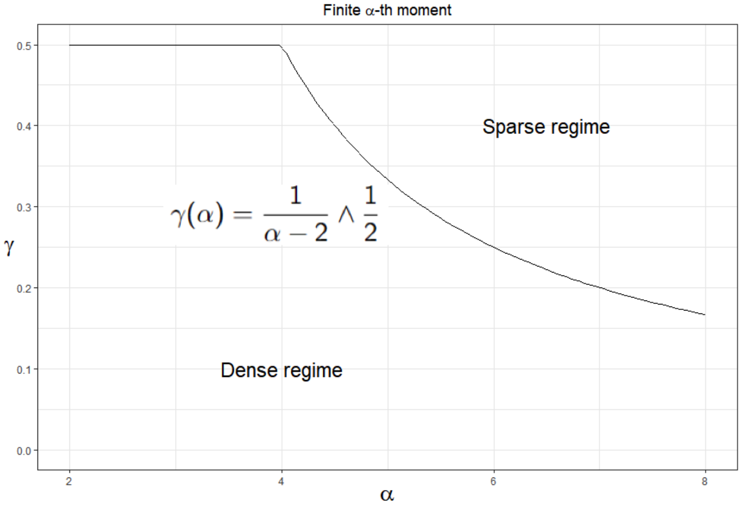
Note: $\omega_1 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)}\right\}}$ and $\omega_2 = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)} \text{ and } \alpha \geq 4\right\}}$.

Three messages

In $\mathcal{P}_{\alpha,K}$, each entry of the noise matrix E has finite α-th moment. For high-dimensional mean change point testing problem:

1. When $\alpha \leq 4$, the sparse regime disappears.
2. Median-of-means-type statistics are effective in handling heavy-tailed data.
3. When $\alpha \geq 4$, in the sparse regime, we propose a computationally efficient test that achieves minimax optimality.

Transition boundary under $\mathcal{P}_{\alpha,K}$



$$s^*_{\mathcal{P}} = p^{\frac{1}{2}} p^{-\gamma(\alpha)}$$

Finite moment $\mathcal{P}_{\alpha,K}$ – dense lower bound

- Consider the dense regime

$$s \geq p^{\frac{1}{2} - \left(\frac{1}{\alpha-2} \wedge \frac{1}{2}\right)}.$$

For $\alpha \geq 2$, we show that $\mathcal{R}_{\mathcal{P}}(\rho) \geq 1/2$ whenever

$$\rho^2 \lesssim p^{(2/\alpha)\vee(1/2)} (\log\log(8n))^{\omega/2} + \log\log(8n),$$

with $\omega = \mathbb{1}_{\left\{s > \sqrt{p \log\log(8n)}\right\} \cap \{\alpha \geq 4\}}$.

- When $\alpha \leq 4$, the dense regime becomes $s \gtrsim 1$, i.e. no sparse regime.

Detour: Median-of-means (MoM)

Let $X_1, \dots, X_n \in \mathbb{R}$ be i.i.d random variables with mean μ and variance σ^2 and consider the following **median-of-means estimator**

$$\hat{\mu}_{\mathrm{MoM}} = \text{median}\left(\frac{1}{m}\sum_{i=1}^{m} X_i, \dots, \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i \right).$$

Let $\delta \in (0,1), k = \lceil 8\log(1/\delta) \rceil$ and $m = n/k$. Then w.p. at least $1-\delta$,

$$|\hat{\mu}_{\mathrm{MoM}} - \mu| \leq \sigma \sqrt{\frac{32\log(1/\delta)}{n}}.$$

Detour: Median-of-means (MoM)

Let $X_1, \ldots, X_n \in \mathbb{R}$ be i.i.d random variables with mean μ and variance σ^2 and consider the following **median-of-means estimator**

$$\hat{\mu}_{\mathrm{MoM}} = \text{median}\left(\frac{1}{m}\sum_{i=1}^{m} X_i, \ldots, \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i \right).$$

Let $\delta \in (0,1), k = \lceil 8\log(1/\delta) \rceil$ and $m = n/k$. Then w.p. at least $1-\delta$,

$$|\hat{\mu}_{\mathrm{MoM}} - \mu| \le \sigma\sqrt{\frac{32\log(1/\delta)}{n}}.$$

- 'sub-Gaussian' property.
- δ is an input to the estimator, through k (number of groups).

Detour: Median-of-means (MoM)

Let $X_1, \dots, X_n \in \mathbb{R}$ be i.i.d random variables with mean μ and variance σ^2 and consider the following **median-of-means estimator**

$$\hat{\mu}_{\mathrm{MoM}} = \text{median}\left(\frac{1}{m}\sum_{i=1}^{m} X_i, \dots, \frac{1}{m}\sum_{i=(k-1)m+1}^{km} X_i\right).$$

Let $\delta \in (0,1), k = \lceil 8\log(1/\delta) \rceil$ and $m = n/k$. Then w.p. at least $1-\delta$,

$$|\hat{\mu}_{\mathrm{MoM}} - \mu| \leq \sigma\sqrt{\frac{32\log(1/\delta)}{n}}.$$

- ‘sub-Gaussian’ property.
- δ is an input to the estimator, through k (number of groups).
- For a given δ, the result is only possible when n is at least $8\log(1/\delta)$.
- Equivalently, for n fixed, δ needs to be larger than $\exp(-n/8)$.

Finite moment $\mathcal{P}_{\alpha,K}$ – dense

- (In sub-Weibull dense...) For $t \in \mathcal{T}$, we compute

$$Y_t := \frac{\sum_{i=1}^{t} X_i - \sum_{i=1}^{t} X_{n+1-i}}{\sqrt{2t}} \in \mathbb{R}^p.$$

and

$$A_t := \sum_{j=1}^{p} \big(Y_t^2(j) - 1\big).$$

Finite moment $\mathcal{P}_{\alpha,K}$ – dense

- For $i \leq n/2$, denote

$$Z_i := (X_i - X_{n-i+1})/\sqrt{2}.$$

- For $t \in \mathcal{T}$, split $\{Z_1, \ldots, Z_t\}$ into G_t groups of equal size

$$\mathcal{Z}_{t,1}, \mathcal{Z}_{t,2}, \ldots, \mathcal{Z}_{t,G_t}.$$

  Each group contains t/G_t elements.

Finite moment $\mathcal{P}_{\alpha,K}$ – dense

- For $i \leq n/2$, denote

$$Z_i := (X_i - X_{n-i+1})/\sqrt{2}.$$

- For $t \in \mathcal{T}$, split $\{Z_1, \ldots, Z_t\}$ into G_t groups of equal size

$$\mathcal{Z}_{t,1}, \mathcal{Z}_{t,2}, \ldots, \mathcal{Z}_{t,G_t}.$$

  Each group contains t/G_t elements.

- Set $V_{t,g} \in \mathbb{R}^p$ with

$$V_{t,g}(j) := \overline{Z}_{t,g}^2(j) - \frac{G_t}{t},$$

  where $\overline{Z}_{t,g}$ is the sample mean of the g-th group $\mathcal{Z}_{t,g}$.

- Median-of-means type statistic:

$$A_t^{\mathrm{MoM}} := t \cdot \mathrm{median}\left(\sum_{j=1}^{p} V_{t,1}(j), \sum_{j=1}^{p} V_{t,2}(j), \ldots, \sum_{j=1}^{p} V_{t,G_t}(j)\right).$$

Finite moment $\mathcal{P}_{\alpha,K}$ – dense

Test:

$$\phi_{\mathcal{P},\text{dense}} := \mathbb{1}_{\left\{\max_{t\in\mathcal{T}} A_t^{\text{MoM}}/r_t > 1\right\}}.$$

Theorem. Assume $\alpha \geq 2$. Choose $G_t = \min\{t, \Delta\}$ and $r_t = Cp^{(2/\alpha)\vee(1/2)}G_t$, with $\Delta = 8\log\log(8n)$. Then $\mathcal{R}_{\mathcal{P}}(\rho, \phi_{\mathcal{P},\text{dense}}) \leq 1/2$ as long as

$$\rho^2 \gtrsim p^{(2/\alpha)\vee(1/2)} \log\log(8n).$$

- When $t \in \mathcal{T} \cap \{t \leq \Delta\}$, MoM simply becomes median.
- When $t \in \mathcal{T} \cap \{t > \Delta\}$, number of groups G_t is at most $\log\log(8n)$.

Finite moment $\mathcal{P}_{\alpha,K}$ – sparse

- In the sparse regime, using the MoM approach with thresholding and sample splitting yields slightly sub-optimal rate.

Finite moment $\mathcal{P}_{\alpha,K}$ – sparse

- In the sparse regime, using the MoM approach with thresholding and sample splitting yields slightly sub-optimal rate.
- Alternative strategy: for each $t \in \mathcal{T} \cap \{t > \tilde{\Delta}\}$, directly apply a robust sparse mean estimator $\hat{\mu}(\cdot)$ to

$$\Big\{ Z_i = (X_i - X_{n+1-i})/\sqrt{2}, \quad i = 1, \ldots, t \Big\}$$

and use $A_t^{\mathrm{RSM}} := t\|\hat{\mu}\|_2^2$ as the test statistic.

- One example of such estimator $\hat{\mu}(\cdot)$ is given in Prasad et al. (2019):

$$\inf_{\mu \in \mathbb{R}^p : \|\mu\|_0 \le s} \sup_{u \in \mathcal{N}_{2s}^{1/2}(\mathcal{S}^{p-1})} \Big| u^\top \mu - \mathrm{1DRobust}(\{u^\top W_i\}_{i=1}^n, \eta/(6ep/s)^s) \Big|,$$

Finite moment $\mathcal{P}_{\alpha,K}$ – sparse

- In the sparse regime, using the MoM approach with thresholding and sample splitting yields slightly sub-optimal rate.
- Alternative strategy: for each $t \in \mathcal{T} \cap \{t > \tilde{\Delta}\}$, directly apply a robust sparse mean estimator $\hat{\mu}(\cdot)$ to

$$\left\{ Z_i = (X_i - X_{n+1-i})/\sqrt{2}, \quad i = 1, \ldots, t \right\}$$

and use $A_t^{\mathrm{RSM}} := t\|\hat{\mu}\|_2^2$ as the test statistic.

- One example of such estimator $\hat{\mu}(\cdot)$ is given in Prasad et al. (2019):

$$\inf_{\mu \in \mathbb{R}^p : \|\mu\|_0 \le s} \sup_{u \in \mathcal{N}_{2s}^{1/2}(\mathcal{S}^{p-1})} \left| u^\top \mu - \mathrm{1DRobust}(\{u^\top W_i\}_{i=1}^n, \eta/(6ep/s)^s) \right|,$$

  – 1DRobust: a univariate robust mean est. (e.g. MoM, trimmed mean).
  – High computational complexity: $|\mathcal{N}_{2s}^{1/2}(\mathcal{S}^{p-1})| \le (6ep/s)^s$.

Finite moment $\mathcal{P}_{\alpha,K}$ – sparse

- We can construct a test $\phi^{\mathrm{RSM}}_{\mathcal{P},\mathrm{sparse}}$ (non-robust when $t \in \mathcal{T} \cap \{t \leq \tilde{\Delta}\}$) that satisfies $\mathcal{R}_{\mathcal{P}}\big(\rho, \phi^{\mathrm{RSM}}_{\mathcal{P},\mathrm{sparse}}\big) \leq 1/2$ as long as

$$\rho^2 \gtrsim s(p/s)^{2/\alpha} + \log\log(8n).$$

Minimax optimal!

Finite moment $\mathcal{P}_{\alpha,K}$ – sparse

- We can construct a test $\phi^{\mathrm{RSM}}_{\mathcal{P},\mathrm{sparse}}$ (non-robust when $t \in \mathcal{T} \cap \{t \leq \tilde{\Delta}\}$) that satisfies $\mathcal{R}_{\mathcal{P}}\big(\rho, \phi^{\mathrm{RSM}}_{\mathcal{P},\mathrm{sparse}}\big) \leq 1/2$ as long as

$$\rho^2 \gtrsim s(p/s)^{2/\alpha} + \log\log(8n).$$

  Minimax optimal!

- To overcome the computation issue, we only use this test when $p \leq \log^{\alpha-2}(\log(8n))$, and use MoM + thresholding + sample splitting otherwise. Best of both worlds!

Summary

- Quantify the costs of heavy-tailedness on the fundamental difficulty of change point testing problems for high-dimensional data.

Summary

- Quantify the costs of heavy-tailedness on the fundamental difficulty of change point testing problems for high-dimensional data.
- Under $\mathcal{G}_{\alpha,K}$, a CUSUM-type test achieves minimax testing rate up to $\sqrt{\log\log(8n)}$.

Summary

- Quantify the costs of heavy-tailedness on the fundamental difficulty of change point testing problems for high-dimensional data.
- Under $\mathcal{G}_{\alpha,K}$, a CUSUM-type test achieves minimax testing rate up to $\sqrt{\log\log(8n)}$.
- Under $\mathcal{P}_{\alpha,K}$, a median-of-means-type test achieves near-optimal testing rate in both dense and sparse regimes.
- In the sparse regime, a computationally efficient procedure can achieve exact optimality.
- Phase transition at $\alpha = 4$ for $\mathcal{P}_{\alpha,K}$ – no sparse regime when $2 \leq \alpha \leq 4$.

Reference

Li, M.*, Chen, Y.*, Wang, T. and Yu, Y. (2023) Robust mean change point testing in high-dimensional data with heavy tails. *arXiv preprint*, arXiv: 2305.18987.

Thank you!