Subgaussian estimators of the mean

Matthieu Lerasle

CNRS, Nice

with L. Devroye, G. Lugosi and R. Oliveira

Subgaussian estimators

Let $\mathcal{P}$ be a family of probability measures P on $\mathbb{R}$ with respective mean E_P and variance σ_P^2. Let $n \in \mathbb{N} \setminus \{0\}$, $L > 0$ and $x_n > 0$.

Subgaussian estimators

Let $\mathcal{P}$ be a family of probability measures P on $\mathbb{R}$ with respective mean E_P and variance σ_P^2. Let $n \in \mathbb{N} \setminus \{0\}$, $L > 0$ and $x_n > 0$.

Definition

1. A single-x L-subgaussian estimator on $(\mathcal{P}, x_n)$ is a map $\widehat{E} : \mathbb{R}^n \times [0, x_n] \to \mathbb{R}$ such that, for any $P \in \mathcal{P}$, if $X_1^n = (X_1, \dots, X_n) \sim P^{\otimes n}$,

$$\forall x \leq x_n, \qquad \mathbb{P}\left(\left| \widehat{E}(X_1^n, x) - E_P \right| > L\sigma_P \sqrt{\frac{1+x}{n}} \right) \leq e^{-x} \quad .$$

   Hereafter, we denote by $\widehat{E}_x(.) = \widehat{E}(., x)$.

Subgaussian estimators

Let $\mathcal{P}$ be a family of probability measures P on $\mathbb{R}$ with respective mean E_P and variance σ_P^2. Let $n \in \mathbb{N} \setminus \{0\}$, $L > 0$ and $x_n > 0$.

Definition

1. A single-x L-subgaussian estimator on $(\mathcal{P}, x_n)$ is a map $\widehat{E} : \mathbb{R}^n \times [0, x_n] \to \mathbb{R}$ such that, for any $P \in \mathcal{P}$, if $X_1^n = (X_1, \ldots, X_n) \sim P^{\otimes n}$,

$$\forall x \leq x_n, \qquad \mathbb{P}\left(\left| \widehat{E}(X_1^n, x) - E_P \right| > L\sigma_P \sqrt{\frac{1+x}{n}} \right) \leq e^{-x} \quad .$$

   Hereafter, we denote by $\widehat{E}_x(.) = \widehat{E}(., x)$.

2. A multiple-x L-subgaussian estimator on $(\mathcal{P}, x_n)$ is a map $\widehat{E} : \mathbb{R}^n \to \mathbb{R}$ such that, for any $P \in \mathcal{P}$, if $X_1^n = (X_1, \ldots, X_n) \sim P^{\otimes n}$,

$$\forall x \leq x_n, \qquad \mathbb{P}\left(\left| \widehat{E}(X_1^n) - E_P \right| > L\sigma_P \sqrt{\frac{1+x}{n}} \right) \leq e^{-x} \quad .$$

Position of the problem

Our goal is to find, for some large classes $\mathcal{P} \subset \mathcal{P}_2$, the largest x_n for which there exists subgaussian estimators over $(\mathcal{P}, x_n)$.

Position of the problem

Our goal is to find, for some large classes $\mathcal{P} \subset \mathcal{P}_2$, the largest x_n for which there exists subgaussian estimators over $(\mathcal{P}, x_n)$.

Example of classes:

$$\mathcal{P}_2^{\sigma} = \{ P \in \mathcal{P}_2, \text{ s. t. } \sigma_P = \sigma \} \quad .$$

$$\mathcal{P}_2^{\leq\sigma} = \{ P \in \mathcal{P}_2, \text{ s. t. } \sigma_P \leq \sigma \} \quad .$$

$$\mathcal{P}_4^{\leq\kappa} = \left\{ P \in \mathcal{P}_4, \text{ s. t. } \frac{\left(P|X - E_P|^4 \right)^{1/4}}{\sigma_P} \leq \kappa \right\} \quad .$$

Position of the problem

Our goal is to find, for some large classes $\mathcal{P} \subset \mathcal{P}_2$, the largest x_n for which there exists subgaussian estimators over $(\mathcal{P}, x_n)$.

Example of classes:

$$\mathcal{P}_2^{\sigma} = \left\{ P \in \mathcal{P}_2, \text{ s. t. } \sigma_P = \sigma \right\} \quad .$$

$$\mathcal{P}_2^{\leq\sigma} = \left\{ P \in \mathcal{P}_2, \text{ s. t. } \sigma_P \leq \sigma \right\} \quad .$$

$$\mathcal{P}_4^{\leq\kappa} = \left\{ P \in \mathcal{P}_4, \text{ s. t. } \frac{\left(P|X - E_P|^4 \right)^{1/4}}{\sigma_P} \leq \kappa \right\} \quad .$$

We provide a general method to *build* subgaussian estimators.

We also present a generic strategy to *prove* some impossibility results.

Behavior of the empirical mean

1. The central limit theorem ensures that, for any fixed x, the empirical mean $\widehat{E}_{\mathrm{emp}}$ is asymptotically subgaussian over $\mathcal{P}_2$, with $L = \sqrt{2}$.

Behavior of the empirical mean

1. The central limit theorem ensures that, for any fixed x, the empirical mean $\widehat{E}_{\mathrm{emp}}$ is asymptotically subgaussian over $\mathcal{P}_2$, with $L = \sqrt{2}$.
2. Catoni (2012) proved that $\widehat{E}_{\mathrm{emp}}$ is $\sqrt{2} + o(1)$-subgaussian over $(\mathcal{P}_4^{\leq \kappa}, x_n)$ for any x_n such that $e^{x_n} = o(\log n)$.

Behavior of the empirical mean

1. The central limit theorem ensures that, for any fixed x, the empirical mean $\widehat{E}_{\mathrm{emp}}$ is asymptotically subgaussian over $\mathcal{P}_2$, with $L = \sqrt{2}$.
2. Catoni (2012) proved that $\widehat{E}_{\mathrm{emp}}$ is $\sqrt{2} + o(1)$-subgaussian over $(\mathcal{P}_4^{\leq \kappa}, x_n)$ for any x_n such that $e^{x_n} = o(\log n)$.
3. Bernstein's inequality proves that $\widehat{E}_{\mathrm{emp}}$ is $\sqrt{2} + o(1)$-subgaussian over $(\mathcal{P}_{\mathrm{exp}}, x_n)$ if $\mathcal{P}_{\mathrm{exp}}$ is a set of probability measures having exponential moments, for any $x_n = o(n)$.

Behavior of the empirical mean

1. The central limit theorem ensures that, for any fixed x, the empirical mean $\widehat{E}_{\mathrm{emp}}$ is asymptotically subgaussian over $\mathcal{P}_2$, with $L = \sqrt{2}$.
2. Catoni (2012) proved that $\widehat{E}_{\mathrm{emp}}$ is $\sqrt{2} + o(1)$-subgaussian over $(\mathcal{P}_4^{\leq \kappa}, x_n)$ for any x_n such that $e^{x_n} = o(\log n)$.
3. Bernstein's inequality proves that $\widehat{E}_{\mathrm{emp}}$ is $\sqrt{2} + o(1)$-subgaussian over $(\mathcal{P}_{\mathrm{exp}}, x_n)$ if $\mathcal{P}_{\mathrm{exp}}$ is a set of probability measures having exponential moments, for any $x_n = o(n)$.
4. The rate $o(n)$ for x_n in the last result cannot be improved unless P has subgaussian tails by the Gärtner-Ellis Theorem.

The median of mean principle

Suppose to simplify that x divides n and let $B_1, \ldots, B_x$ denotes a partition of $\{1, \ldots, n\}$ into sets B_i with cardinality $|B_i| = n/x$.

The median of mean principle

Suppose to simplify that x divides n and let $B_1, \ldots, B_x$ denotes a partition of $\{1, \ldots, n\}$ into sets B_i with cardinality $|B_i| = n/x$.

For any i, let $Y_i = \frac{x}{n} \sum_{j \in B_i} X_j$, (the vector of means).

The median of mean principle

Suppose to simplify that x divides n and let $B_1, \dots, B_x$ denotes a partition of $\{1, \dots, n\}$ into sets B_i with cardinality $|B_i| = n/x$.

For any i, let $Y_i = \frac{x}{n} \sum_{j \in B_i} X_j$, (the vector of means).

Define $\widehat{E}_x = \mathrm{median}(Y_i)$ (the median of means).

The median of mean principle

Suppose to simplify that x divides n and let $B_1, \dots, B_x$ denotes a partition of $\{1, \dots, n\}$ into sets B_i with cardinality $|B_i| = n/x$.

For any i, let $Y_i = \frac{x}{n} \sum_{j \in B_i} X_j$, (the vector of means).

Define $\widehat{E}_x = \mathrm{median}(Y_i)$ (the median of means).

Lemma

$$\mathbb{P}\left(\left| \widehat{E}_x - E_P \right| > 2e\sigma_P \sqrt{\frac{x}{n}} \right) \leq e^{-x} \ .$$

Proof

By Tchebycheff's inequality

$$\forall i, \qquad \mathbb{P}\left(|Y_i - E_P| > 2e\sigma_P \sqrt{\frac{x}{n}} \right) \leq \frac{1}{(2e)^2} \quad .$$

Proof

By Tchebycheff's inequality

$$\forall i, \qquad \mathbb{P}\left(|Y_i - E_P| > 2e\sigma_P\sqrt{\frac{x}{n}} \right) \leq \frac{1}{(2e)^2} \quad .$$

Now by definition

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{E}_x - E_P\right| > 2e\sigma_P\sqrt{\frac{x}{n}} \right) \\
&\leq \mathbb{P}\left(\#\left\{ i, \text{ s. t. } |Y_i - E_P| > 2e\sigma_P\sqrt{\frac{x}{n}} \right\} \geq \frac{x}{2} \right) \\
&\leq \mathbb{P}\left(\mathrm{Bin}\left(x, \frac{1}{(2e)^2} \right) \geq \frac{x}{2} \right) \\
&\leq \sum_{k \geq x/2} \binom{x}{k} \left(\frac{1}{(2e)^2} \right)^k \left(1 - \frac{1}{(2e)^2} \right)^{x-k}
\end{aligned}$$

Proof

By Tchebycheff's inequality

$$\forall i, \qquad \mathbb{P}\left(|Y_i - E_P| > 2e\sigma_P \sqrt{\frac{x}{n}} \right) \leq \frac{1}{(2e)^2} \ .$$

Now by definition

$$\begin{aligned}
&\mathbb{P}\left(\left|\widehat{E}_x - E_P\right| > 2e\sigma_P \sqrt{\frac{x}{n}} \right) \\
&\leq \mathbb{P}\left(\# \left\{ i, \text{ s. t. } |Y_i - E_P| > 2e\sigma_P \sqrt{\frac{x}{n}} \right\} \geq \frac{x}{2} \right) \\
&\leq \mathbb{P}\left(\mathrm{Bin}\left(x, \frac{1}{(2e)^2} \right) \geq \frac{x}{2} \right) \\
&\leq \sum_{k \geq x/2} \binom{x}{k} \underbrace{\left(\frac{1}{(2e)^2} \right)^k \left(1 - \frac{1}{(2e)^2} \right)^{x-k}}_{\leq \frac{1}{(2e)^x}}
\end{aligned}$$

Proof

By Tchebycheff's inequality

$$\forall i, \qquad \mathbb{P}\left(|Y_i - E_P| > 2e\sigma_P\sqrt{\frac{x}{n}} \right) \leq \frac{1}{(2e)^2} \quad .$$

Now by definition

$$\begin{aligned}
&\mathbb{P}\left(\left| \widehat{E}_x - E_P \right| > 2e\sigma_P\sqrt{\frac{x}{n}} \right) \\
&\leq \mathbb{P}\left(\# \left\{ i, \text{ s. t. } |Y_i - E_P| > 2e\sigma_P\sqrt{\frac{x}{n}} \right\} \geq \frac{x}{2} \right) \\
&\leq \mathbb{P}\left(\mathrm{Bin}\left(x, \frac{1}{(2e)^2} \right) \geq \frac{x}{2} \right) \\
&\leq \underbrace{\sum_{k \geq x/2} \binom{x}{k}}_{\leq 2^x} \underbrace{\left(\frac{1}{(2e)^2} \right)^k \left(1 - \frac{1}{(2e)^2} \right)^{x-k}}_{\leq \frac{1}{(2e)^x}}
\end{aligned}$$

Proof

By Tchebycheff's inequality

$$\forall i, \qquad \mathbb{P}\left(|Y_i - E_P| > 2e\sigma_P\sqrt{\frac{x}{n}} \right) \leq \frac{1}{(2e)^2} \quad .$$

Now by definition

$$\begin{aligned}
&\mathbb{P}\left(\left| \widehat{E}_x - E_P \right| > 2e\sigma_P\sqrt{\frac{x}{n}} \right) \\
&\leq \mathbb{P}\left(\# \left\{ i, \text{ s. t. } |Y_i - E_P| > 2e\sigma_P\sqrt{\frac{x}{n}} \right\} \geq \frac{x}{2} \right) \\
&\leq \mathbb{P}\left(\mathrm{Bin}\left(x, \frac{1}{(2e)^2} \right) \geq \frac{x}{2} \right) \\
&\leq \underbrace{\sum_{k \geq x/2} \binom{x}{k}}_{\leq 2^x} \underbrace{\left(\frac{1}{(2e)^2} \right)^k \left(1 - \frac{1}{(2e)^2} \right)^{x-k}}_{\leq \frac{1}{(2e)^x}} \leq e^{-x} \quad .
\end{aligned}$$

The confidence intervals method

Suppose $\sigma_P^2 = \sigma^2$ is *known*.

The confidence intervals method

Suppose $\sigma_P^2 = \sigma^2$ is *known*.

Using the median of means principle, one can build, for any $k \leq Cn$, a confidence interval $\widehat{I}_k = \left[\widehat{E}_k \pm L\sigma\sqrt{\frac{k}{n}}\right]$ for E_P with confidence level e^{-k}.

The confidence intervals method

Suppose $\sigma_P^2 = \sigma^2$ is *known*.

Using the median of means principle, one can build, for any $k \leq Cn$, a confidence interval $\widehat{I}_k = \left[\widehat{E}_k \pm L\sigma\sqrt{\frac{k}{n}}\right]$ for E_P with confidence level e^{-k}.

Let

$$\widehat{k} = \inf\left\{k \leq Cn, \text{ s. t. } \bigcap_{j=k}^{Cn} \widehat{I}_j \neq \emptyset\right\} .$$

The confidence intervals method

Suppose $\sigma_P^2 = \sigma^2$ is *known*.

Using the median of means principle, one can build, for any $k \leq Cn$, a confidence interval $\widehat{I}_k = \left[\widehat{E}_k \pm L\sigma\sqrt{\frac{k}{n}} \right]$ for E_P with confidence level e^{-k}.

Let

$$\widehat{k} = \inf\left\{ k \leq Cn, \text{ s. t. } \bigcap_{j=k}^{Cn} \widehat{I}_j \neq \emptyset \right\} \quad .$$

$\cap_{j=\widehat{k}}^{Cn} \widehat{I}_j$ is a non-empty closed interval, let $\widehat{E}$ denote its midpoint.

The confidence intervals method

Suppose $\sigma_P^2 = \sigma^2$ is *known*.

Using the median of means principle, one can build, for any $k \leq Cn$, a confidence interval $\widehat{I}_k = \left[\widehat{E}_k \pm L\sigma\sqrt{\frac{k}{n}} \right]$ for E_P with confidence level e^{-k}.

Let

$$\widehat{k} = \inf \left\{ k \leq Cn, \text{ s. t. } \bigcap_{j=k}^{Cn} \widehat{I}_j \neq \emptyset \right\} \quad .$$

$\cap_{j=\widehat{k}}^{Cn} \widehat{I}_j$ is a non-empty closed interval, let $\widehat{E}$ denote its midpoint.

Theorem

$\widehat{E}$ is a multiple-x $2\sqrt{2}L$-subgaussian estimator for $(\mathcal{P}_2^{\sigma^2}, Cn - 2)$.

Proof

Let $x \in [0, Cn - 2]$ and let $k \leq \lfloor x \rfloor + 2$.

Proof

Let $x \in [0, Cn - 2]$ and let $k \leq \lfloor x \rfloor + 2$. By construction,

$$\mathbb{P}\left(E_P \notin \bigcap_{j=k}^{Cn} \widehat{I}_j\right) \leq \sum_{j=k}^{Cn} \mathbb{P}\left(E_P \notin \widehat{I}_j\right) \leq e^{1-k} \leq e^{-x} \quad .$$

Proof

Let $x \in [0, Cn - 2]$ and let $k \leq \lfloor x \rfloor + 2$. By construction,

$$\mathbb{P}\left(E_P \notin \bigcap_{j=k}^{Cn} \widehat{I}_j \right) \leq \sum_{j=k}^{Cn} \mathbb{P}\left(E_P \notin \widehat{I}_j \right) \leq e^{1-k} \leq e^{-x} \quad .$$

When $E_P \in \bigcap_{j=k}^{Cn} \widehat{I}_j$, then $E_P \in \widehat{I}_k$ and $\bigcap_{j=k}^{Cn} \widehat{I}_j \neq \emptyset$ so $\widehat{k} \leq k$.

Proof

Let $x \in [0, Cn-2]$ and let $k \leq \lfloor x \rfloor + 2$. By construction,

$$\mathbb{P}\left(E_P \notin \bigcap_{j=k}^{Cn} \widehat{I}_j\right) \leq \sum_{j=k}^{Cn} \mathbb{P}\left(E_P \notin \widehat{I}_j\right) \leq e^{1-k} \leq e^{-x} \quad .$$

When $E_P \in \bigcap_{j=k}^{Cn} \widehat{I}_j$, then $E_P \in \widehat{I}_k$ and $\bigcap_{j=k}^{Cn} \widehat{I}_j \neq \emptyset$ so $\widehat{k} \leq k$.

When $\widehat{k} \leq k$, $\widehat{E} \in \bigcap_{j=k}^{Cn} \widehat{I}_j$, so $\widehat{E} \in \widehat{I}_k$.

Proof

Let $x \in [0, Cn-2]$ and let $k \leq \lfloor x \rfloor + 2$. By construction,

$$\mathbb{P}\left(E_P \notin \bigcap_{j=k}^{Cn} \widehat{I}_j \right) \leq \sum_{j=k}^{Cn} \mathbb{P}\left(E_P \notin \widehat{I}_j \right) \leq e^{1-k} \leq e^{-x} \quad .$$

When $E_P \in \bigcap_{j=k}^{Cn} \widehat{I}_j$, then $E_P \in \widehat{I}_k$ and $\bigcap_{j=k}^{Cn} \widehat{I}_j \neq \emptyset$ so $\widehat{k} \leq k$.

When $\widehat{k} \leq k$, $\widehat{E} \in \bigcap_{j=k}^{Cn} \widehat{I}_j$, so $\widehat{E} \in \widehat{I}_k$.

When both E_P and $\widehat{E}$ belong to $\widehat{I}_k$,

$$|\widehat{E} - E_P| \leq 2L\sigma\sqrt{\frac{k}{n}} \leq 2L\sigma\sqrt{\frac{x+2}{n}} \leq 2L\sqrt{2}\sigma\sqrt{\frac{1+x}{n}} \quad .$$

Other classes $\mathcal{P}$

The method of confidence intervals can be used to build multiple-x subgaussian intervals

1. over the class $\mathcal{P}_2^{\mathrm{sym}} = \{\, P \in \mathcal{P}_2, \text{ s. t. } X \sim P \implies 2E_P - X \sim P \,\}$,

Other classes $\mathcal{P}$

The method of confidence intervals can be used to build multiple-x subgaussian intervals

1. over the class $\mathcal{P}_2^{\mathrm{sym}} = \{ P \in \mathcal{P}_2, \text{ s. t. } X \sim P \implies 2E_P - X \sim P \}$,
2. over the classes $\mathcal{P}_\alpha^{\le \eta} = \{ P \in \mathcal{P}_\alpha, \text{ s. t. } \mathbb{E}_P \left[|X - E_P|^\alpha \right] \le (\eta \sigma_P)^\alpha \}$, for $\alpha = 3, 4$ and $\eta > 0$.

Laplace distributions

The method of confidence intervals allows to build subgaussian estimators with $x_n \geq cn$ assuming only a *known* variance of the data. These clearly outperform the empirical mean.

Laplace distributions

The method of confidence intervals allows to build subgaussian estimators with $x_n \geq cn$ assuming only a *known* variance of the data. These clearly outperform the empirical mean.

We show now that the order n of x_n is the best possible.

Laplace distributions

The method of confidence intervals allows to build subgaussian estimators with $x_n \geq cn$ assuming only a *known* variance of the data. These clearly outperform the empirical mean.

We show now that the order n of x_n is the best possible.

For any $\lambda \in \mathbb{R}$, let La_λ denote the distribution on $\mathbb{R}$ with density

$$f_\lambda(x) = \frac{e^{-|x-\lambda|}}{2} \quad .$$

Laplace distributions

The method of confidence intervals allows to build subgaussian estimators with $x_n \geq cn$ assuming only a *known* variance of the data. These clearly outperform the empirical mean.

We show now that the order n of x_n is the best possible.

For any $\lambda \in \mathbb{R}$, let La_λ denote the distribution on $\mathbb{R}$ with density

$$f_\lambda(x) = \frac{e^{-|x-\lambda|}}{2} \quad .$$

This distribution satisfies $E_{\mathrm{La}_\lambda} = \lambda$ and $\sigma^2_{\mathrm{La}_\lambda} = 2$. Let $\mathcal{P}_{\mathrm{La}} = \{\mathrm{La}_\lambda,\ \lambda \in \mathbb{R}\}$.

Laplace distributions

The method of confidence intervals allows to build subgaussian estimators with $x_n \geq cn$ assuming only a *known* variance of the data. These clearly outperform the empirical mean.

We show now that the order n of x_n is the best possible.

For any $\lambda \in \mathbb{R}$, let La_λ denote the distribution on $\mathbb{R}$ with density

$$f_\lambda(x) = \frac{e^{-|x-\lambda|}}{2} \quad .$$

This distribution satisfies $E_{\mathrm{La}_\lambda} = \lambda$ and $\sigma^2_{\mathrm{La}_\lambda} = 2$. Let $\mathcal{P}_{\mathrm{La}} = \{\mathrm{La}_\lambda,\ \lambda \in \mathbb{R}\}$.

Theorem

Let $L \geq \sqrt{2}$ and $C = 9L^2$. There doesn't exist single-x L-subgaussian estimators over $(\mathcal{P}_{\mathrm{La}}, Cn)$.

Proof

By contradiction let $x = 9L^2n - 1$ and let $\widehat{E}_x$ be a single-x, L-subgaussian estimator. Let $\lambda = 2L\sqrt{2\frac{1+x}{n}}$, $X_1^n \sim \mathrm{La}_0^{\otimes n}$, $Y_1^n \sim \mathrm{La}_\lambda^{\otimes n}$.

Proof

By contradiction let $x = 9L^2 n - 1$ and let $\widehat{E}_x$ be a single-x, L-subgaussian estimator. Let $\lambda = 2L\sqrt{2\frac{1+x}{n}}$, $X_1^n \sim \mathrm{La}_0^{\otimes n}$, $Y_1^n \sim \mathrm{La}_\lambda^{\otimes n}$. By the triangular inequality

$$\forall (x_1, \ldots, x_n) \in \mathbb{R}^n, \qquad \prod_{i=1}^n f_0(x_i) \geq e^{-\lambda n} \prod_{i=1}^n f_\lambda(x_i) \enspace .$$

Proof

By contradiction let $x = 9L^2 n - 1$ and let $\widehat{E}_x$ be a single-x, L-subgaussian estimator. Let $\lambda = 2L\sqrt{2\frac{1+x}{n}}$, $X_1^n \sim \mathrm{La}_0^{\otimes n}$, $Y_1^n \sim \mathrm{La}_\lambda^{\otimes n}$. By the triangular inequality

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \qquad \prod_{i=1}^n f_0(x_i) \geq e^{-\lambda n} \prod_{i=1}^n f_\lambda(x_i) \ .$$

Therefore,

$$\mathbb{P}\left(|\widehat{E}_x(X_1^n) - E_P| > \frac{\lambda}{2} \right) \geq e^{-\lambda n} \mathbb{P}\left(|\widehat{E}_x(Y_1^n) - E_P| > \frac{\lambda}{2} \right) \ .$$

Proof

By contradiction let $x = 9L^2n - 1$ and let $\widehat{E}_x$ be a single-x, L-subgaussian estimator. Let $\lambda = 2L\sqrt{2\frac{1+x}{n}}$, $X_1^n \sim \mathrm{La}_0^{\otimes n}$, $Y_1^n \sim \mathrm{La}_\lambda^{\otimes n}$. By the triangular inequality

$$\forall (x_1, \ldots, x_n) \in \mathbb{R}^n, \qquad \prod_{i=1}^n f_0(x_i) \geq e^{-\lambda n} \prod_{i=1}^n f_\lambda(x_i) \ .$$

Therefore,

$$\mathbb{P}\left(|\widehat{E}_x(X_1^n) - E_P| > \frac{\lambda}{2} \right) \geq e^{-\lambda n} \mathbb{P}\left(|\widehat{E}_x(Y_1^n) - E_P| > \frac{\lambda}{2} \right) \ .$$

Thus, by definition of λ and the subgaussian property.

$$e^{-x} \geq e^{-\lambda n}(1 - e^{-x}) \ .$$

Proof

By contradiction let $x = 9L^2 n - 1$ and let $\widehat{E}_x$ be a single-x, L-subgaussian estimator. Let $\lambda = 2L\sqrt{2\frac{1+x}{n}}$, $X_1^n \sim \mathrm{La}_0^{\otimes n}$, $Y_1^n \sim \mathrm{La}_\lambda^{\otimes n}$. By the triangular inequality

$$\forall (x_1, \dots, x_n) \in \mathbb{R}^n, \qquad \prod_{i=1}^{n} f_0(x_i) \geq e^{-\lambda n} \prod_{i=1}^{n} f_\lambda(x_i) \ .$$

Therefore,

$$\mathbb{P}\left(|\widehat{E}_x(X_1^n) - E_P| > \frac{\lambda}{2} \right) \geq e^{-\lambda n} \mathbb{P}\left(|\widehat{E}_x(Y_1^n) - E_P| > \frac{\lambda}{2} \right) \ .$$

Thus, by definition of λ and the subgaussian property.

$$e^{-x} \geq e^{-\lambda n}(1 - e^{-x}) \ .$$

Again by definition of λ and x, this implies $e^{(9-6\sqrt{2})L^2 n} \leq e^{1+\log 2}$ which is wrong for any $n \geq 2$ since $L \geq \sqrt{2}$.

multiple-x and single-x are different notions

Theorem

For any $L \geq \sqrt{2}$ and any $x_n \to \infty$, there doesn't exist multiple-x L-subgaussian estimators over $(\mathcal{P}_2, x_n)$.

The proof relies on the same kind of minimax arguments, but we used the class $\mathcal{P}_{\mathrm{Po}}$ of Poisson distributions instead of the class $\mathcal{P}_{\mathrm{La}}$.

multiple-x and single-x are different notions

Theorem

For any $L \geq \sqrt{2}$ and any $x_n \to \infty$, there doesn't exist multiple-x L-subgaussian estimators over $(\mathcal{P}_2, x_n)$.

The proof relies on the same kind of minimax arguments, but we used the class $\mathcal{P}_{\mathrm{Po}}$ of Poisson distributions instead of the class $\mathcal{P}_{\mathrm{La}}$.

Consequences

1. Multiple-x and single-x are different notions (by the median of means principle, there *exists* single-x L-subgaussian estimators over $(\mathcal{P}_2, cn)$).

multiple-x and single-x are different notions

Theorem

For any $L \geq \sqrt{2}$ and any $x_n \to \infty$, there doesn't exist multiple-x L-subgaussian estimators over $(\mathcal{P}_2, x_n)$.

The proof relies on the same kind of minimax arguments, but we used the class $\mathcal{P}_{\mathrm{Po}}$ of Poisson distributions instead of the class $\mathcal{P}_{\mathrm{La}}$.

Consequences

1. Multiple-x and single-x are different notions (by the median of means principle, there *exists* single-x L-subgaussian estimators over $(\mathcal{P}_2, cn)$).
2. One cannot derive multiple-x from single-x subgaussian estimators.

Getting optimal L

Catoni (2012) proved that

1. $L = \sqrt{2}$ is optimal on any class $\mathcal{P} \supset (\mathcal{N}(m, \sigma^2))_{m \in \mathbb{R}}$.

Getting optimal L

Catoni (2012) proved that

1. $L = \sqrt{2}$ is optimal on any class $\mathcal{P} \supset (\mathcal{N}(m, \sigma^2))_{m \in \mathbb{R}}$.
2. there exists single-x $(\sqrt{2} + o(1))$-subgaussian estimators over $(\mathcal{P}_2, \frac{n}{2} - 1)$.

Getting optimal L

Catoni (2012) proved that

1. $L = \sqrt{2}$ is optimal on any class $\mathcal{P} \supset (\mathcal{N}(m, \sigma^2))_{m\in\mathbb{R}}$.
2. there exists single-x $(\sqrt{2} + o(1))$-subgaussian estimators over $(\mathcal{P}_2, \frac{n}{2} - 1)$.

Theorem

For any $x_n = o\left(\left(\frac{n}{\sqrt{\kappa}}\right)^{2/3}\right)$, there exists multiple-$x$ $(\sqrt{2} + o(1))$-subgaussian estimators over $(\mathcal{P}_4^{\leq\kappa}, x_n)$.

I: Truncation of the data

For any (E, R), define

$$\Psi_{E,R}(x) = \mu + \left(\frac{R}{|x - E|} \wedge 1 \right) (x - E) \ .$$

I: Truncation of the data

For any (E, R), define

$$\Psi_{E,R}(x) = \mu + \left(\frac{R}{|x - E|} \wedge 1 \right) (x - E) \ .$$

Lemma

If $E = E_P$ and $R = \sigma_P \sqrt{\frac{n}{x_n}}$, then, for any $x \leq x_n$,

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \Psi_{E,R}(X_i) - E_P \right| > (\sqrt{2} + o(1)) \sigma_P \sqrt{\frac{t}{n}} \right) \leq 1 - 2e^{-t} \ .$$

I: Truncation of the data

For any (E, R), define

$$\Psi_{E,R}(x) = \mu + \left(\frac{R}{|x-E|} \wedge 1 \right)(x - E) \ .$$

Lemma

If $E = E_P$ and $R = \sigma_P \sqrt{\frac{n}{x_n}}$, then, for any $x \le x_n$,

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \Psi_{E,R}(X_i) - E_P \right| > (\sqrt{2} + o(1))\sigma_P \sqrt{\frac{t}{n}} \right) \le 1 - 2e^{-t} \ .$$

Proof: exponential Markov's inequality.

II: an insensitivity argument

Let $\mathcal{R} = \left\{ (E, R), \text{ s. t. } |E - E_P| \leq \varepsilon_1 \sigma_P, \ \left| R - \sigma_P \sqrt{\frac{n}{2x_n}} \right| \leq \varepsilon_2 \sigma_P \right\}$ and let

$$\Delta_{E,R} = \frac{1}{n} \sum_{i=1}^{n} (\Psi_{E,R}(X_i) - \Psi_{E_P, \sigma_P \sqrt{\frac{n}{2x_n}}}(X_i)) \ .$$

II: an insensitivity argument

Let $\mathcal{R} = \left\{ (E, R), \text{ s. t. } |E - E_P| \leq \varepsilon_1 \sigma_P, \ \left| R - \sigma_P \sqrt{\frac{n}{2x_n}} \right| \leq \varepsilon_2 \sigma_P \right\}$ and let

$$\Delta_{E,R} = \frac{1}{n} \sum_{i=1}^{n} (\Psi_{E,R}(X_i) - \Psi_{E_P, \sigma_P \sqrt{\frac{n}{2x_n}}}(X_i)) \ .$$

Lemma

For any $\varepsilon_1 = o(1)$ and $\varepsilon_2 = O(1)$,

$$\mathbb{P}\left(\forall (E, R) \in \mathcal{R}, \qquad |\Delta_{E,R}| \leq o(1) \sigma_P \sqrt{\frac{t}{n}} \right) \geq 1 - e^{-t} \ .$$

II: an insensitivity argument

Let $\mathcal{R} = \left\{ (E,R),\ \text{s. t. } |E - E_P| \le \varepsilon_1 \sigma_P,\ \left| R - \sigma_P \sqrt{\frac{n}{2x_n}} \right| \le \varepsilon_2 \sigma_P \right\}$ and let

$$\Delta_{E,R} = \frac{1}{n} \sum_{i=1}^{n} (\Psi_{E,R}(X_i) - \Psi_{E_P,\sigma_P \sqrt{\frac{n}{2x_n}}}(X_i)) \ .$$

Lemma

For any $\varepsilon_1 = o(1)$ *and* $\varepsilon_2 = O(1)$,

$$\mathbb{P}\left(\forall (E,R) \in \mathcal{R}, \qquad |\Delta_{E,R}| \le o(1) \sigma_P \sqrt{\frac{t}{n}} \right) \ge 1 - e^{-t} \ .$$

Proof: Chaining argument+Bernstein's inequality.

III: Median of means principles

Let $\widehat{E}_{x_n}, \widehat{\sigma}^2_{x_n}$ denote a median of means estimator of E_P and a median of "U-statistics" estimator of σ^2_P, built with $x_n + 1$ blocks.

III: Median of means principles

Let $\widehat{E}_{x_n}, \widehat{\sigma}^2_{x_n}$ denote a median of means estimator of E_P and a median of "*U*-statistics" estimator of σ_P^2, built with $x_n + 1$ blocks.

Lemma

$$\mathbb{P}\left(\left|\widehat{E}_{x_n} - E_P\right| > L\sigma_P\sqrt{\frac{x_n}{n}} \text{ and } \sqrt{\frac{n}{2x_n}}\left|\widehat{\sigma}_{x_n} - \sigma_P\right| \leq \frac{\kappa}{2} \right) \geq 1 - e^{-x_n} \quad .$$

Some references

[1] Catoni (2012). Challenging the empirical mean and empirical variance: a deviation study. Annales de l'IHP P&S.

[2] Hsu and Sabato (2014). Heavy-tailed regression with a generalized median-of-means. ICML.

[3] Nemirovsky and Yudin (1983). Problem complexity and method efficiency in optimization. Wiley interscience.

Thanks