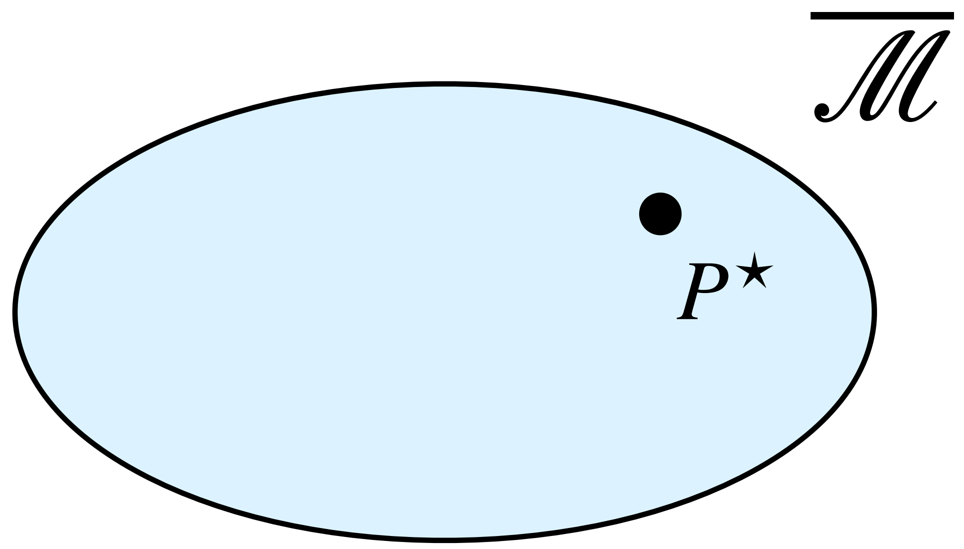
Robust density estimation in TV under a shape constraint

Hélène Halconruy and Guillaume Maillard

Joint work with Y. Baraud

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N°811017

Motivation (1)



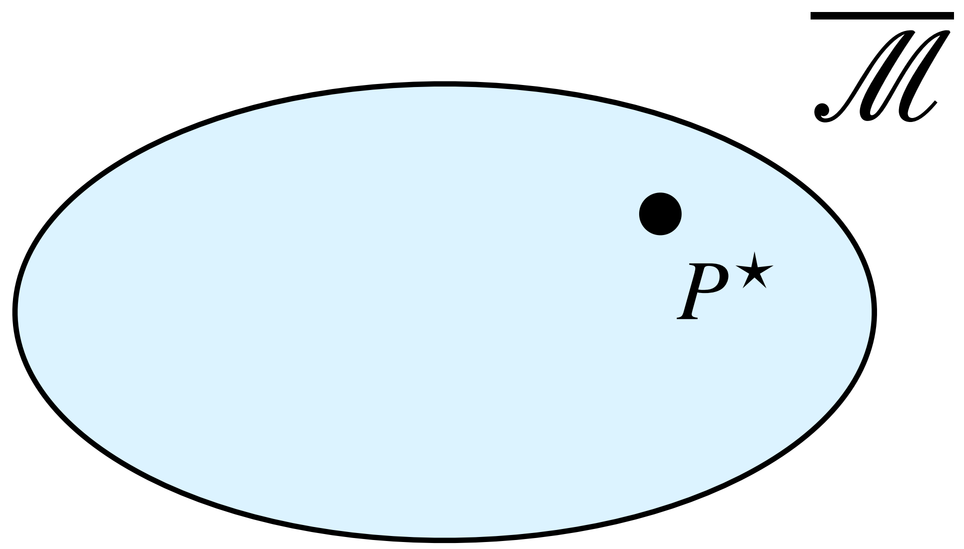
Presumed situation $X_1, \ldots, , X_n$ i.i.d. with law $P^\star$ and density f.

Objective Estimate f. Density on the line that satisfies a shape constraint :

- Monotone on a half-line
- Convex/concave monotone on an interval
- log-concave on the line

[ideal] : $X_1, \ldots, , X_n$ i.i.d. $+ f$ satisfies the shape constraint

Motivation (1)



Presumed situation $X_1, \ldots, , X_n$ i.i.d. with law $P^\star$ and density f.

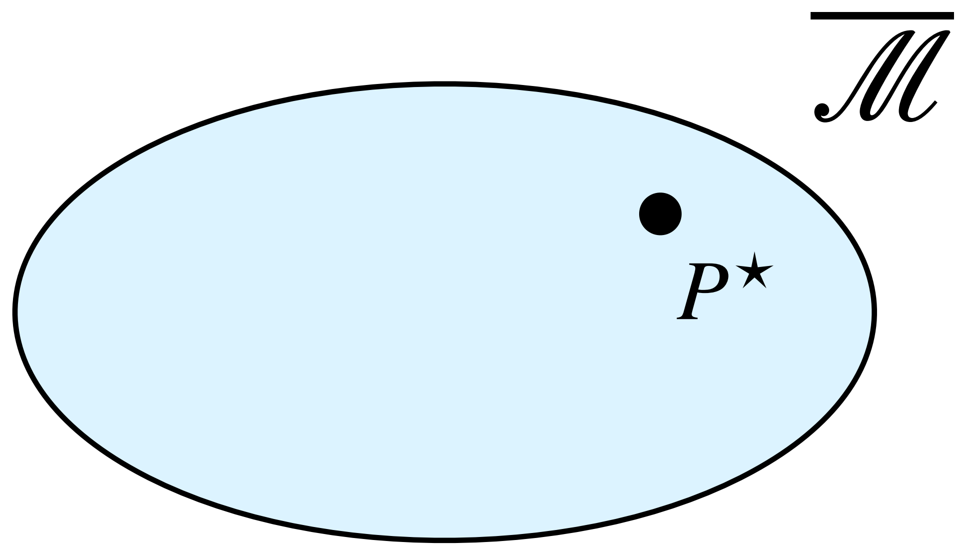
Objective Estimate f. Density on the line that satisfies a shape constraint :

- Monotone on a half-line
- Convex/concave monotone on an interval
- log-concave on the line

[ideal] : $X_1, \ldots, , X_n$ i.i.d. $+ f$ satisfies the shape constraint

Question Behaviour of your estimator if **[ideal]**? When moving away from **[ideal]**?

Motivation (1)



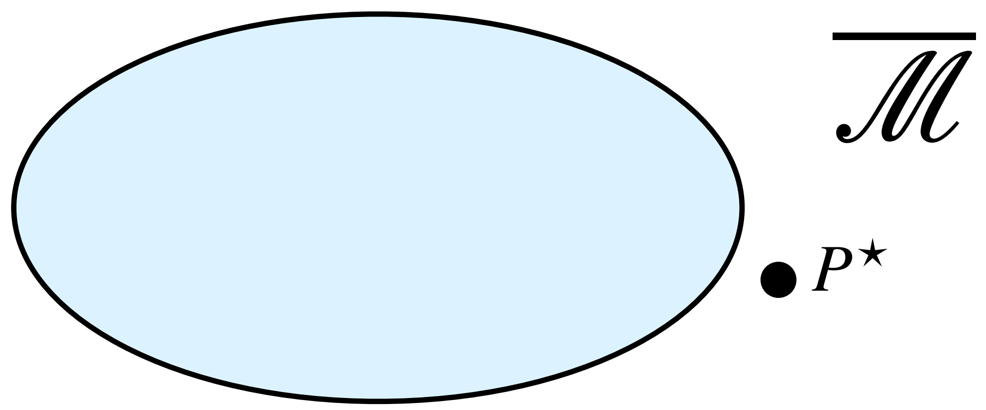
Presumed situation $X_1, \ldots, , X_n$ i.i.d. with law $P^\star$ and density f.

Objective Estimate f. Density on the line that satisfies a shape constraint :

- Monotone on a half-line
- Convex/concave monotone on an interval
- log-concave on the line

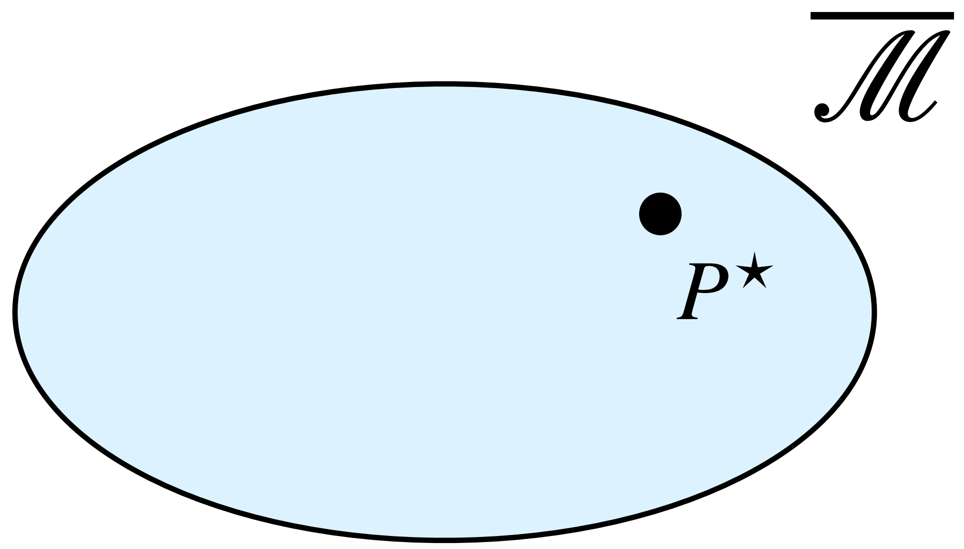
[ideal] : $X_1, \ldots, , X_n$ i.i.d. + ~~f satisfies the shape constraint~~

Question Behaviour of your estimator if **[ideal]**? When moving away from **[ideal]**?



[missp] i.i.d. data but $P^\star \notin \overline{\mathscr{M}}$

Motivation (1)



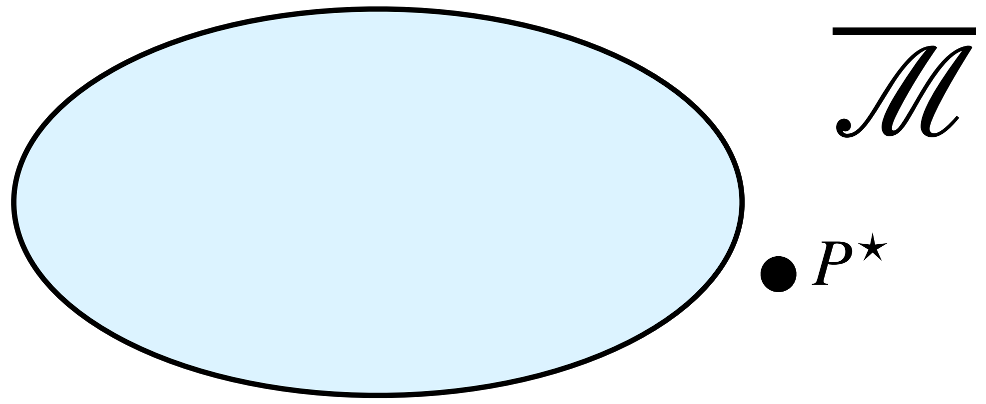
Presumed situation $X_1, \dots, , X_n$ i.i.d. with law $P^\star$ and density f.

Objective Estimate f. Density on the line that satisfies a shape constraint :

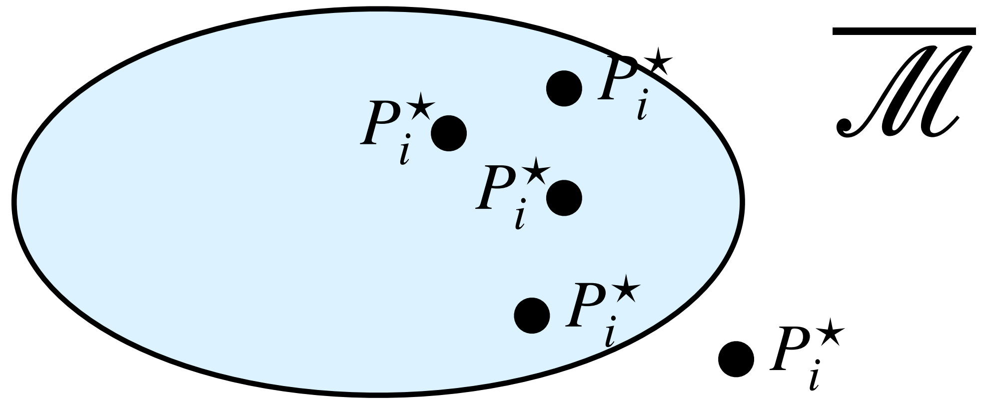
- Monotone on a half-line
- Convex/concave monotone on an interval
- log-concave on the line

[ideal] : ~~$X_1, \dots, , X_n$ i.i.d.~~ + ~~f satisfies the shape constraint~~

Question Behaviour of your estimator if **[ideal]**? When moving away from **[ideal]**?



[missp] i.i.d. data but $P^\star \notin \overline{\mathscr{M}}$

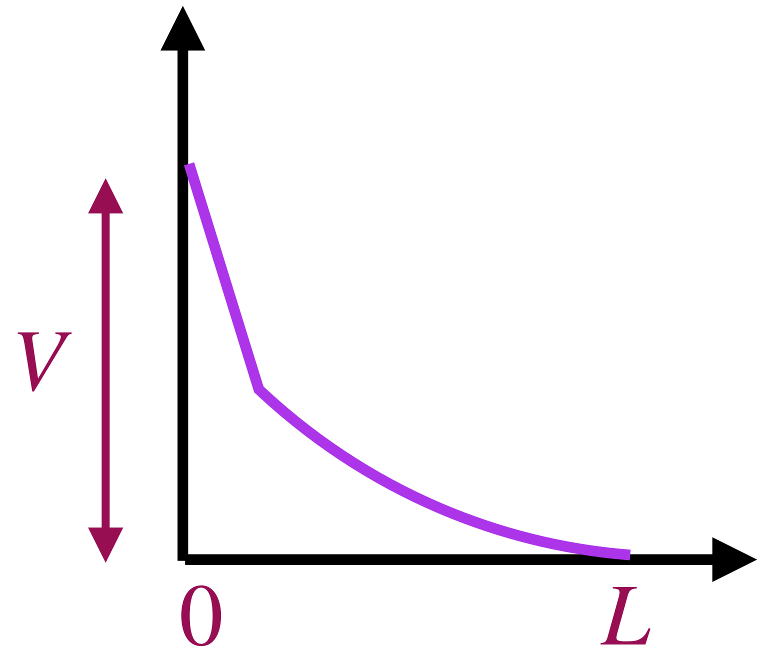


[non iid] independent non i.i.d. data $X_i \sim P_i^\star$

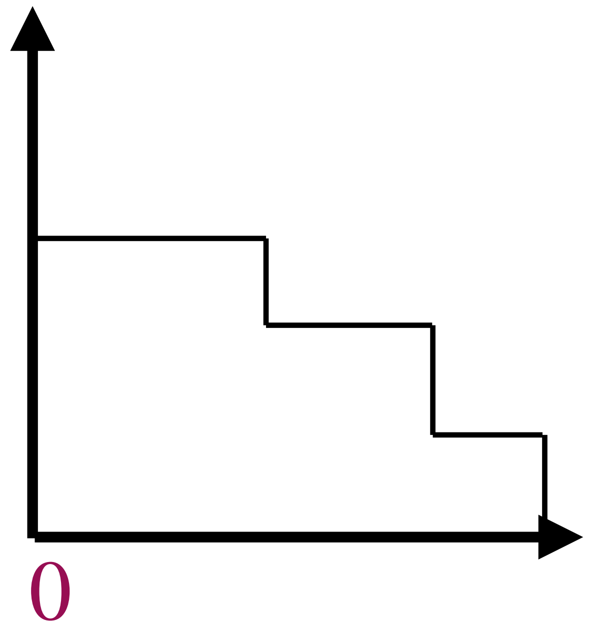
Example Monotonocity on the half line [iid]

f density decreasing on $(0,\infty)$

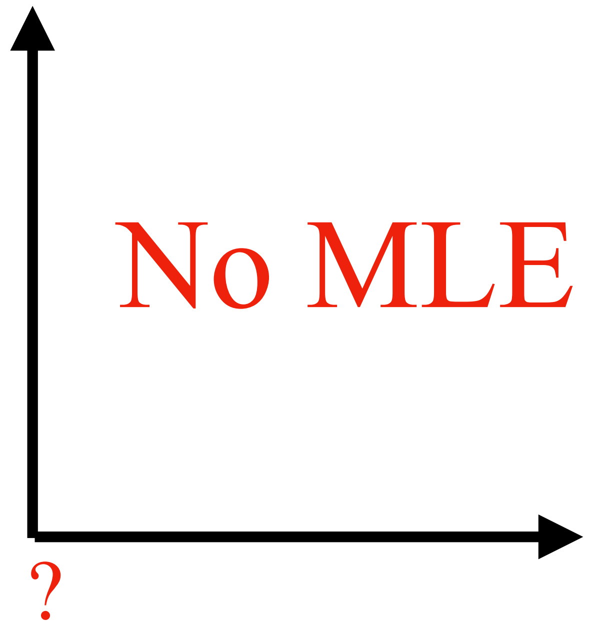
Grenander [1956] = MLE /↘ densities on $(0,\infty)$



+ **[Opt]** $\sim$ minimax over $\mathscr{F}(V, L)$, L and V would be known



+ **[Adapt]** $\left(1/\sqrt{n}\right)$-rate for some piecewise constant densities with $D < \infty$ pieces



– Additional information, not always defined

Aims for the estimation procedure

Objective To build an estimator

[Opt] Global rate of convergence *optimal*
[Adapt] Specific classes *adaptation*
✓ MLE

[Vers] No additional information *versatile*
✗ MLE

General : ℓ-type TV-estimator

$\overline{\mathscr{M}} = \{P = p \cdot \mu, p \in \overline{M}\}$
μ Lebesgue measure

[Baraud, 2021]
[Devroye and Lugosi, 2001]

Total Variation loss $\quad d(P,Q) = \sup_{a \in \mathscr{A}} [P(A) - Q(A)]$

Assumption There exists M (corr. dist. $\mathscr{M}$) countable, dense in $\overline{M}$ (corr. dist. $\overline{\mathscr{M}}$) for $\mathbf{L}^1$

Test statistics
$P = p \cdot \mu$
$Q = q \cdot \mu$

For $P \in \mathscr{M}$,

$$\mathbf{T}(\mathbf{X}, P) = \sup_{Q \in \mathscr{M}} \left[\sum_{i=1}^{n} \mathbf{1}_{\{q>p\}}(X_i) - P(q > p) \right]$$

TV-estimator $\quad \boxed{\widehat{P} = \operatorname{argmin}_{P \in \mathscr{M}} \mathbf{T}(\mathbf{X}, P) \text{ or } \varepsilon\text{-minimizer of } P \mapsto \mathbf{T}(\mathbf{X}, P)}$

Properties of TV-estimator

$\overline{\mathscr{M}} = \{p \cdot \mu, p \in \overline{M}\}$
M countable, dense in $\overline{M}$ for $\mathbf{L}^1$

Definition [Baraud and Birgé, 2016] $D \in \mathbb{N}$

$\overline{p} \in \overline{M}$ is **extremal** in $\overline{M}$ with degree $\leqslant D$ if

$$\left\{\{q > \overline{p}\}, q \in \overline{M} \backslash \{\overline{p}\}\right\} \quad \text{and} \quad \left\{\{q < \overline{p}\}, q \in \overline{M} \backslash \{\overline{p}\}\right\}$$

are both VC with dimension $\leqslant D$. Set of extremal points $\overline{\mathcal{O}} = \bigcup_{D \geqslant 1} \overline{\mathcal{O}}(D)$

Prel. Vers. : result on

$$1/n \sum_{i=1}^{n} d(P_i^\star, \widehat{P}\,) \qquad \overline{P}^\star = n^{-1} \sum_{i=1}^{n} P_i^\star \qquad \overline{P}^\star = P^\star \text{ if } X_1, \ldots, X_n \text{ i.i.d.}$$

Theorem

For all TV-estimator $\widehat{P}$, $\quad \mathbf{E}[d(\overline{P}^\star, \widehat{P}\,)]$

Properties of TV-estimator

$\overline{\mathscr{M}} = \{p \cdot \mu, p \in \overline{M}\}$
M countable, dense in $\overline{M}$ for $\mathbf{L}^1$

Definition [Baraud and Birgé, 2016] $D \in \mathbb{N}$

$\overline{p} \in \overline{M}$ is **extremal** in $\overline{M}$ with degree $\leqslant D$ if

$$\big\{\{q > \overline{p}\}, q \in \overline{M} \backslash \{\overline{p}\}\big\} \quad \text{and} \quad \big\{\{q < \overline{p}\}, q \in \overline{M} \backslash \{\overline{p}\}\big\}$$

are both VC with dimension $\leqslant D$. Set of extremal points $\overline{\mathcal{O}} = \bigcup_{D \geqslant 1} \overline{\mathcal{O}}(D)$

Prel. Vers. : result on

$$1/n \sum_{i=1}^{n} d(P_i^\star, \widehat{P}) \qquad \overline{P}^\star = n^{-1} \sum_{i=1}^{n} P_i^\star \qquad \overline{P}^\star = P^\star \text{ if } X_1, \ldots, X_n \text{ i.i.d.}$$

Theorem

For all TV-estimator $\widehat{P}$, $\mathbf{E}[d(\overline{P}^\star, \widehat{P})]$

Properties of TV-estimator

$\overline{\mathscr{M}} = \{p \cdot \mu, p \in \overline{M}\}$
M countable, dense in $\overline{M}$ for $\mathbf{L}^1$

Definition [Baraud and Birgé, 2016] $D \in \mathbb{N}$

$\overline{p} \in \overline{M}$ is **extremal** in $\overline{M}$ with degree $\leqslant D$ if

$$\left\{\{q > \overline{p}\}, q \in \overline{M} \backslash \{\overline{p}\}\right\} \quad \text{and} \quad \left\{\{q < \overline{p}\}, q \in \overline{M} \backslash \{\overline{p}\}\right\}$$

are both VC with dimension $\leqslant D$. Set of extremal points $\overline{\mathcal{O}} = \bigcup_{D \geqslant 1} \overline{\mathcal{O}}(D)$

Prel. Vers. : result on

$$1/n \sum_{i=1}^{n} d(P_i^\star, \widehat{P}) \qquad \overline{P}^\star = n^{-1} \sum_{i=1}^{n} P_i^\star \qquad \overline{P}^\star = P^\star \text{ if } X_1, \ldots, X_n \text{ i.i.d.}$$

Theorem

For all TV-estimator $\widehat{P}$,

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1} \left[\underbrace{3 \inf_{P \in \overline{\mathcal{O}}(D)} d(\overline{P}^\star, P)}_{\textbf{Approximation error}} + \underbrace{48\sqrt{\frac{D}{n}}}_{\textbf{Estimation error}} \right] + \frac{\varepsilon}{n}$$

Comments

$$\overline{P}^\star = n^{-1}\sum_{i=1}^{n} P_i^\star$$

For all TV-estimator $\widehat{P}$,

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D\geqslant 1}\left[3\inf_{P\in\overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(1) Ancillary result

(AR) For all $P \in \overline{\mathscr{O}}(D)$, $\quad \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(2) Adaptation

Comments

$$\overline{P}^{\star} = n^{-1}\sum_{i=1}^{n} P_i^{\star}$$

For all TV-estimator $\widehat{P}$,

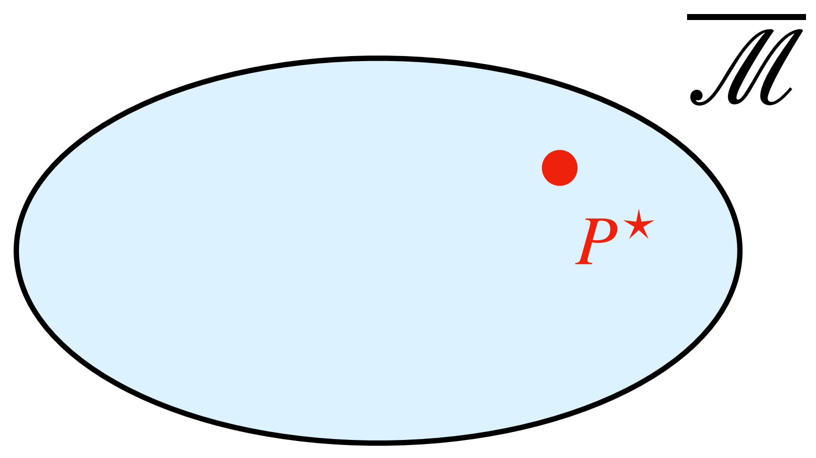
$$\mathbf{E}[d(\overline{P}^{\star}, \widehat{P})] \leqslant \inf_{D \geqslant 1}\left[3 \inf_{P \in \overline{\mathscr{O}}(D)} d(\overline{P}^{\star}, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(1) Ancillary result

(AR) For all $P \in \overline{\mathscr{O}}(D)$, $\quad \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^{\star}, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(2) Adaptation

[ideal]



[Adapt] If **[ideal]** with $\overline{P}^{\star} = P^{\star}$ in $\overline{\mathscr{O}}$

Comments

For all TV-estimator $\widehat{P}$,

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1} \left[3 \inf_{P \in \overline{\mathcal{O}}(D)} \underbrace{\cancel{d(\overline{P}^\star, P)}}_{= 0} + 48 \sqrt{\frac{D}{n}} \right] + \frac{\varepsilon}{n}$$

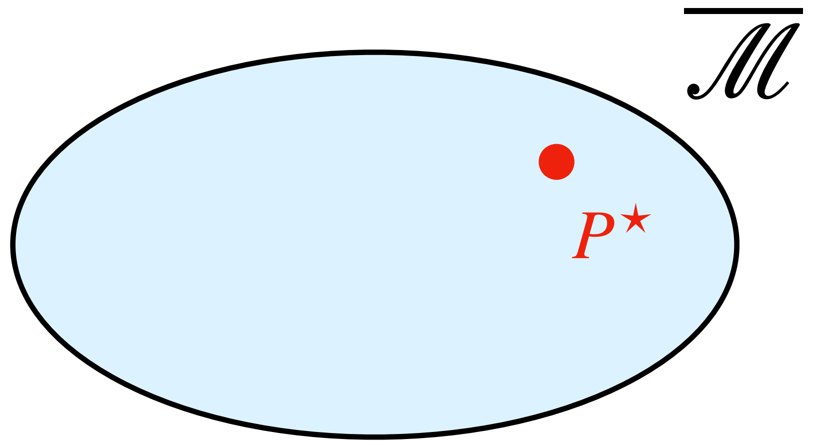
$$\boxed{\overline{P}^\star = n^{-1} \sum_{i=1}^{n} P_i^\star}$$

(1) Ancillary result

$$\textbf{(AR)} \quad \text{For all } P \in \overline{\mathcal{O}}(D), \quad \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

(2) Adaptation

[ideal]



[Adapt] If **[ideal]** with $\overline{P}^\star = P^\star$ in $\overline{\mathcal{O}}$, rate $\sim 1/\sqrt{n}$

Comments

$$\overline{P}^\star = n^{-1}\sum_{i=1}^{n} P_i^\star$$

For all TV-estimator $\widehat{P}$,

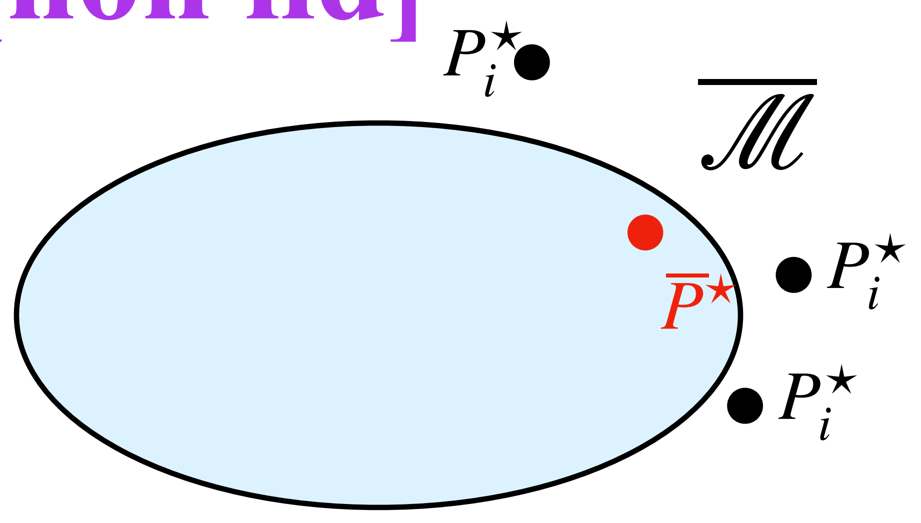
$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D\geqslant 1}\left[3\inf_{P\in\overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(1) Ancillary result

(AR) For all $P \in \overline{\mathscr{O}}(D)$, $\quad \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(2) Adaptation

[non iid]



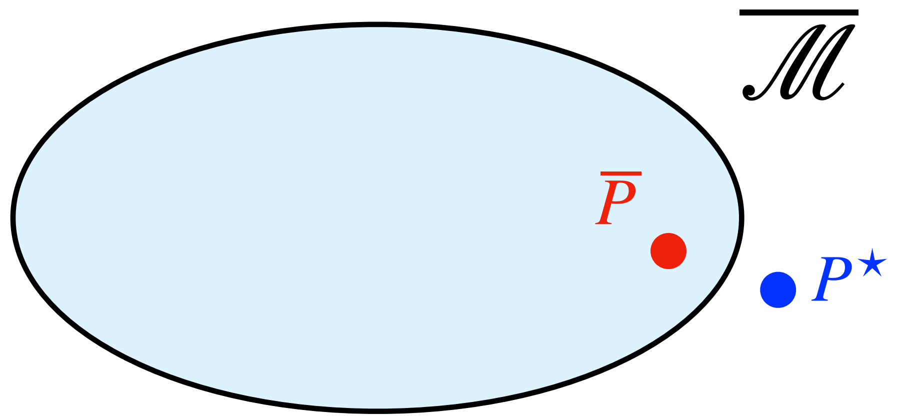
[Adapt] If **[non iid]** with $\overline{P}^\star$ in $\overline{\mathscr{O}}$, rate $\sim 1/\sqrt{n}$

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D\geqslant 1}\left[3 \inf_{P\in\overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(AR) $\forall\, P \in \overline{\mathscr{O}}(D),\ \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(3) Robustness

[A] $X_1, \ldots, , X_n$ i.i.d. + ~~f satisfies the shape constraint~~



[Rob] If **[A]** and $P^\star \notin \overline{\mathscr{M}}$

Misspecification

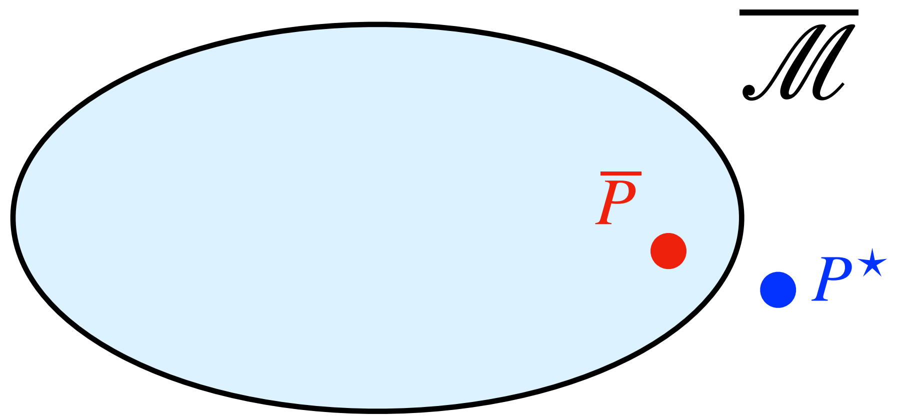
$P^\star$ δ-close to $\overline{P} \in \overline{\mathscr{O}}(D)$ for TV dist.

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D\geqslant 1}\left[3 \inf_{P\in\overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(AR) $\forall\, P \in \overline{\mathscr{O}}(D),\ \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(3) Robustness

[A] $X_1, \ldots, , X_n$ i.i.d. + ~~f satisfies the shape constraint~~



[Rob] If **[A]** and $P^\star \notin \overline{\mathscr{M}}$

Misspecification

$P^\star$ δ-close to $\overline{P} \in \overline{\mathscr{O}}(D)$ for TV dist.

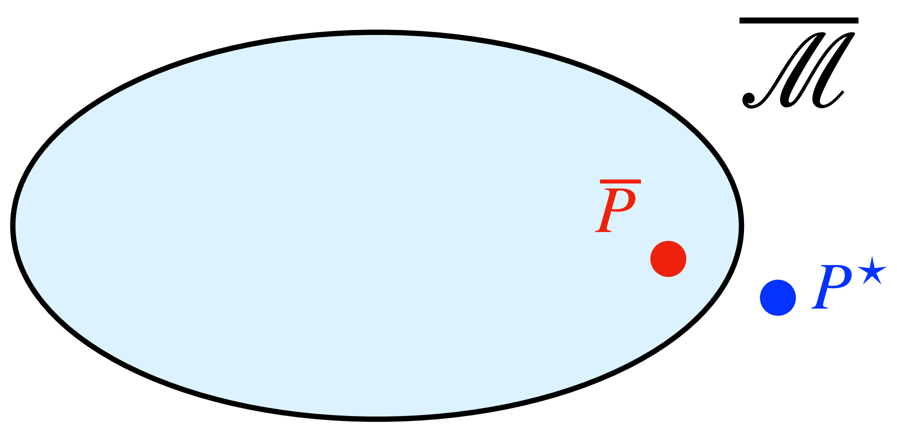
$$\mathbf{E}[d(P^\star, \widehat{P})] \leqslant 3\delta + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

(3) Robustness

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1} \left[3 \inf_{P \in \overline{\mathcal{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} \right] + \frac{\varepsilon}{n}$$

(AR) $\forall\, P \in \overline{\mathcal{O}}(D)$, $\mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

[A] $X_1, \ldots, X_n$ i.i.d. + data contaminated



[Rob] If **[A]** and $P^\star = (1-\alpha)\overline{P} + \alpha R$, $\overline{P} \in \overline{\mathcal{O}}(D)$

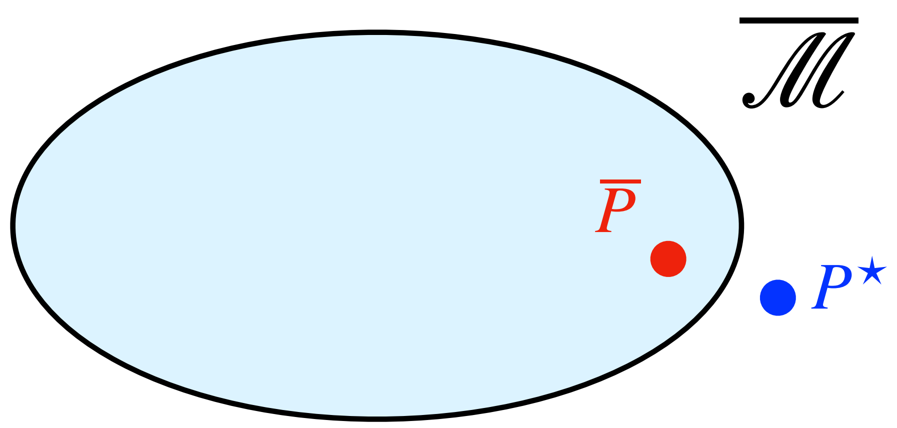
Contamination

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1}\left[3 \inf_{P \in \overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(AR) $\forall\, P \in \overline{\mathscr{O}}(D),\ \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(3) Robustness

[A] $X_1, \ldots, X_n$ i.i.d. + data contaminated



[Rob] If **[A]** and $P^\star = (1-\alpha)\overline{P} + \alpha R,\ \overline{P} \in \overline{\mathscr{O}}(D)$

Contamination

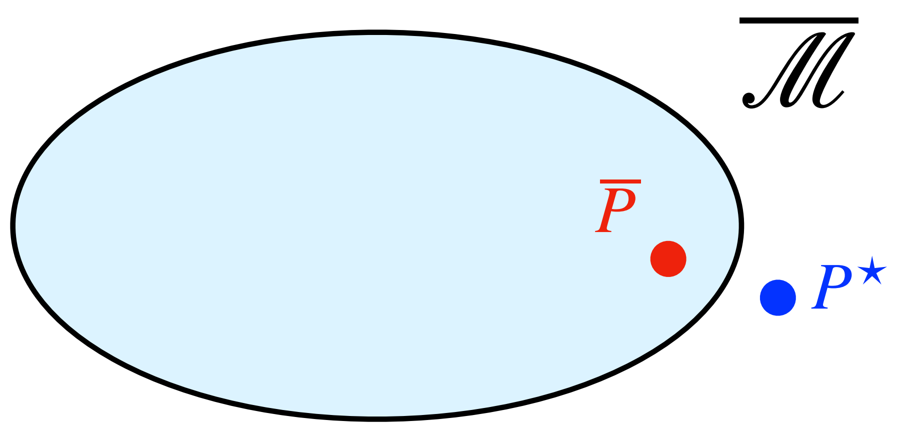
(AR) $\mathbf{E}[d(\overline{P}, \widehat{P})] \leqslant 2\alpha + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(3) Robustness

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1} \left[3 \inf_{P \in \overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

$$\textbf{(AR)}\ \forall\, P \in \overline{\mathscr{O}}(D),\ \ \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

[A] $X_1, \ldots, X_n$ i.i.d. + data contaminated

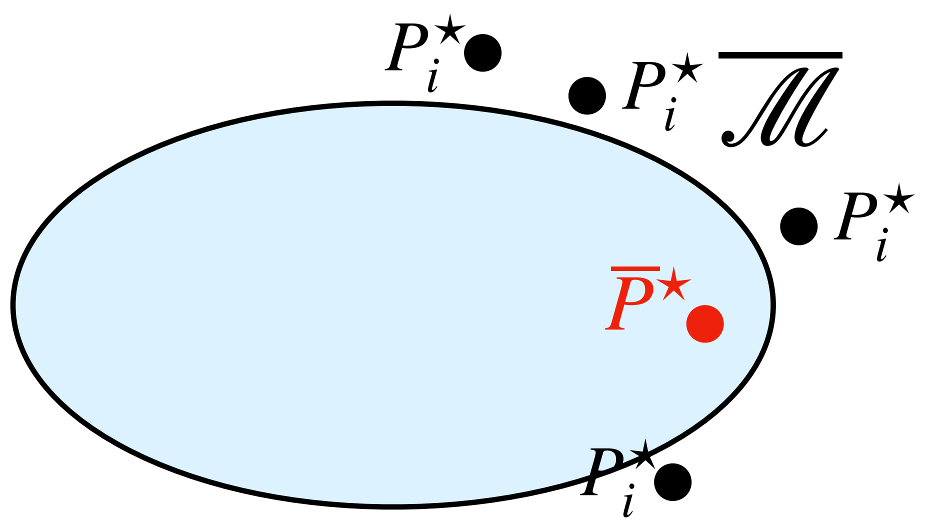


[Rob] If **[A]** and $P^\star = (1-\alpha)\overline{P} + \alpha R$, $\overline{P} \in \overline{\mathscr{O}}(D)$

Contamination

$$\textbf{(AR)} \quad \mathbf{E}[d(\overline{P}, \widehat{P})] \leqslant 2\alpha + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

[B] ~~$X_1, \ldots, X_n$ i.i.d.~~ + ~~f satisfies the shape constraint~~



[Rob] If **[B]** and $\overline{P}^\star$ is α-close to $\overline{P} \in \overline{\mathscr{O}}(D)$ for TV dist.

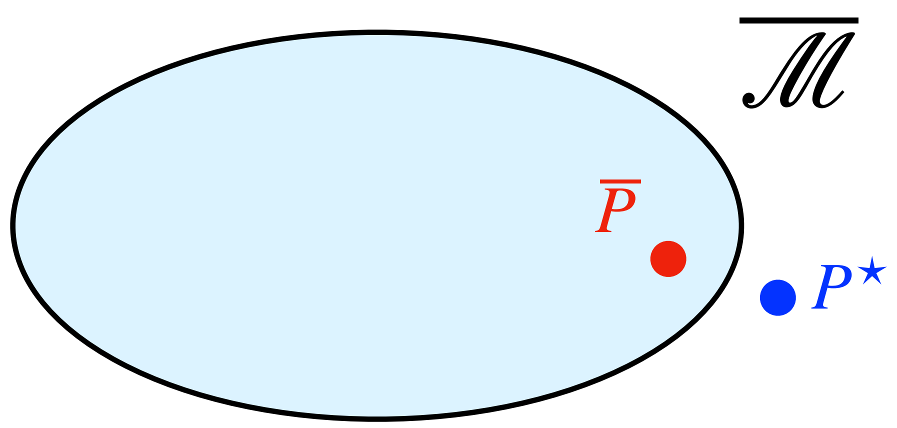
Non-equidistribution

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1}\left[3 \inf_{P \in \overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(AR) $\forall\, P \in \overline{\mathscr{O}}(D),\ \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

(3) Robustness

[A] $X_1, \ldots, , X_n$ i.i.d. + data contaminated

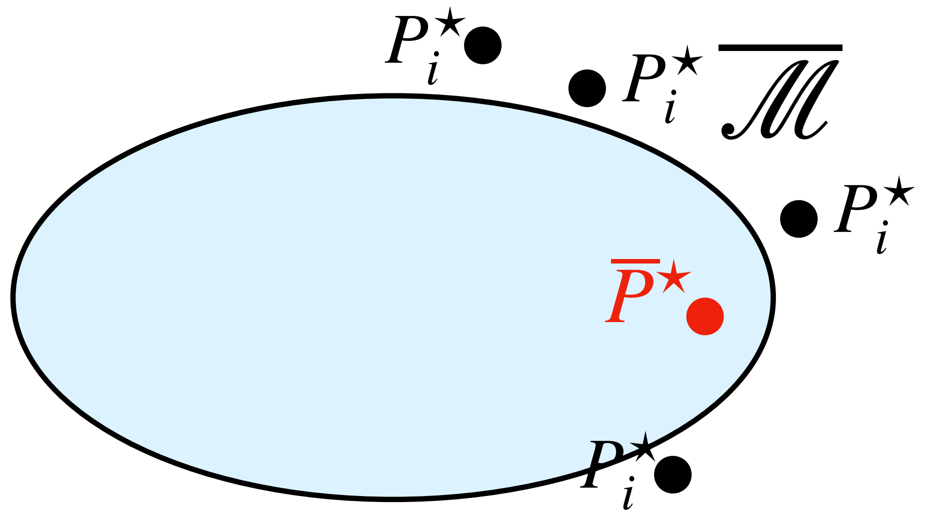


[Rob] If **[A]** and $P^\star = (1-\alpha)\overline{P} + \alpha R$, $\overline{P} \in \overline{\mathscr{O}}(D)$

Contamination

$$\textbf{(AR)} \quad \mathbf{E}[d(\overline{P}, \widehat{P})] \leqslant 2\alpha + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

[B] ~~$X_1, \ldots, , X_n$ i.i.d.~~ + ~~f satisfies the shape constraint~~



[Rob] If **[B]** and $\overline{P}^\star$ is α-close to $\overline{P} \in \overline{\mathscr{O}}(D)$ for TV dist.

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant 3\alpha + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

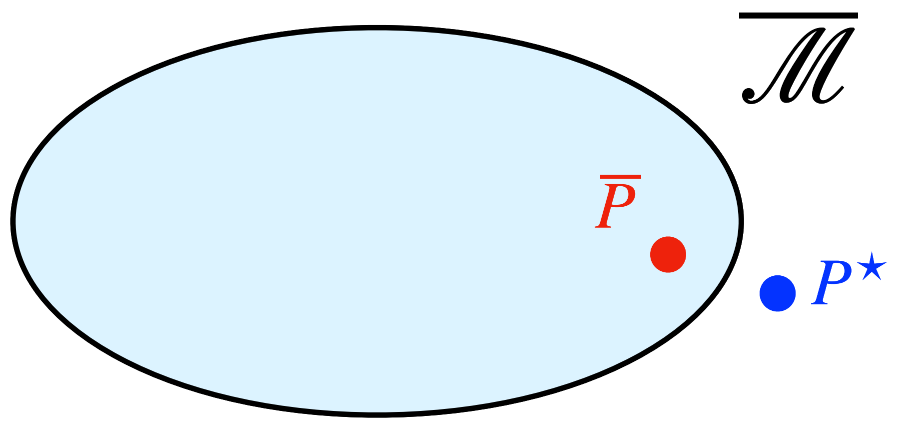
Non-equidistribution

(3) Robustness

$$\mathbf{E}[d(\overline{P}^\star, \widehat{P}\,)] \leqslant \inf_{D\geqslant 1}\left[3 \inf_{P\in\overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}$$

(AR) $\forall\, P \in \overline{\mathscr{O}}(D),\ \mathbf{E}[d(P, \widehat{P}\,)] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$

[A] $X_1, \ldots, X_n$ i.i.d. + data contaminated

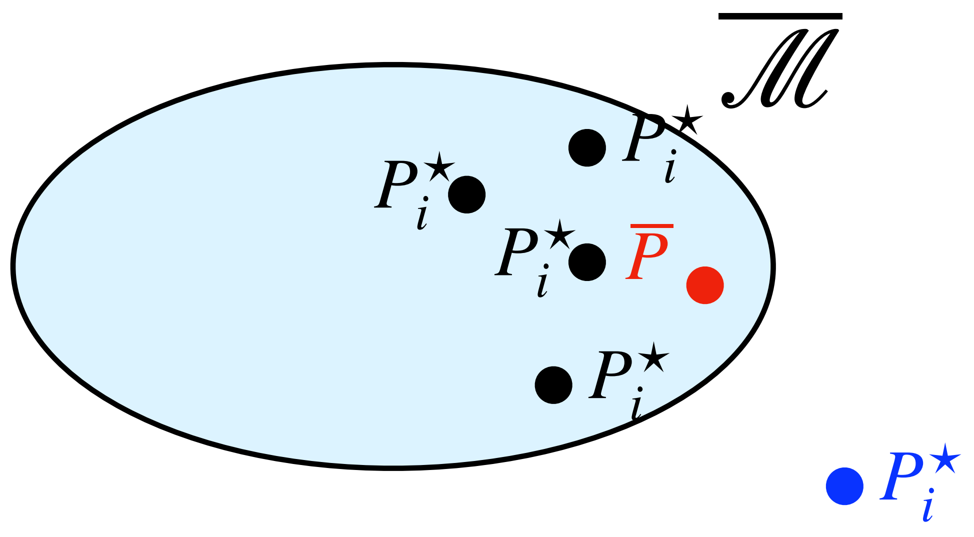


Contamination

✔ **[Rob]** If **[A]** and $P^\star = (1-\alpha)\overline{P} + \alpha\, R$, $\overline{P} \in \overline{\mathscr{O}}(D)$

$$\textbf{(AR)}\quad \mathbf{E}[d(\overline{P}, \widehat{P}\,)] \leqslant 2\alpha + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

[B] ~~$X_1, \ldots, X_n$ i.i.d.~~ + outliers



Outliers

✔ **[Rob]** If **[B]** and

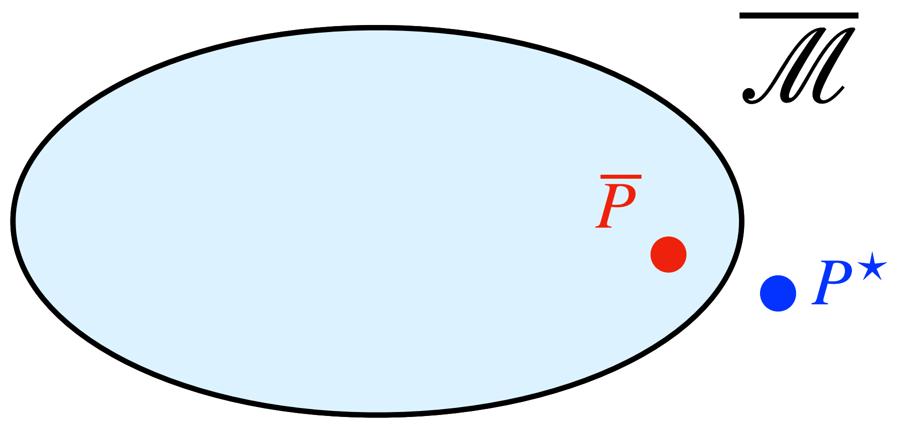
$$P_i^\star = (1-\alpha_i)\,\overline{P} + \alpha_i R_i\ ,\ \overline{P} \in \overline{\mathscr{O}}(D)$$

e.g. outliers $\exists\, S \subset \{1,\ldots,n\}\ :\ \forall i \in S, \alpha_i = 1$ and $\forall i \notin S, \alpha_i = 0$

$$\boxed{\begin{aligned}&\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D\geqslant 1}\left[3 \inf_{P\in\overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}}\right] + \frac{\varepsilon}{n}\\ &\textbf{(AR)}\ \forall P \in \overline{\mathscr{O}}(D),\ \mathbf{E}[d(P, \widehat{P})] \leqslant 2d(\overline{P}^\star, P) + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}\end{aligned}}$$

(3) Robustness

[A] $X_1, \ldots, , X_n$ i.i.d. + data contaminated

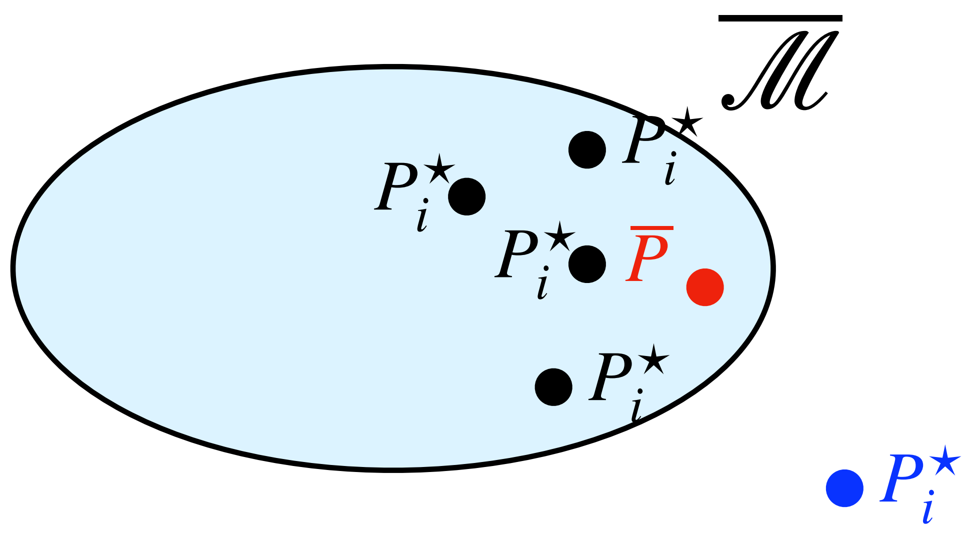


[Rob] If **[A]** and $P^\star = (1-\alpha)\overline{P} + \alpha R$, $\overline{P} \in \overline{\mathscr{O}}(D)$

Contamination

$$\textbf{(AR)}\quad \mathbf{E}[d(\overline{P}, \widehat{P})] \leqslant 2\alpha + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

[B] ~~$X_1, \ldots, X_n$ i.i.d.~~ + outliers



[Rob] If **[B]** and

Outliers

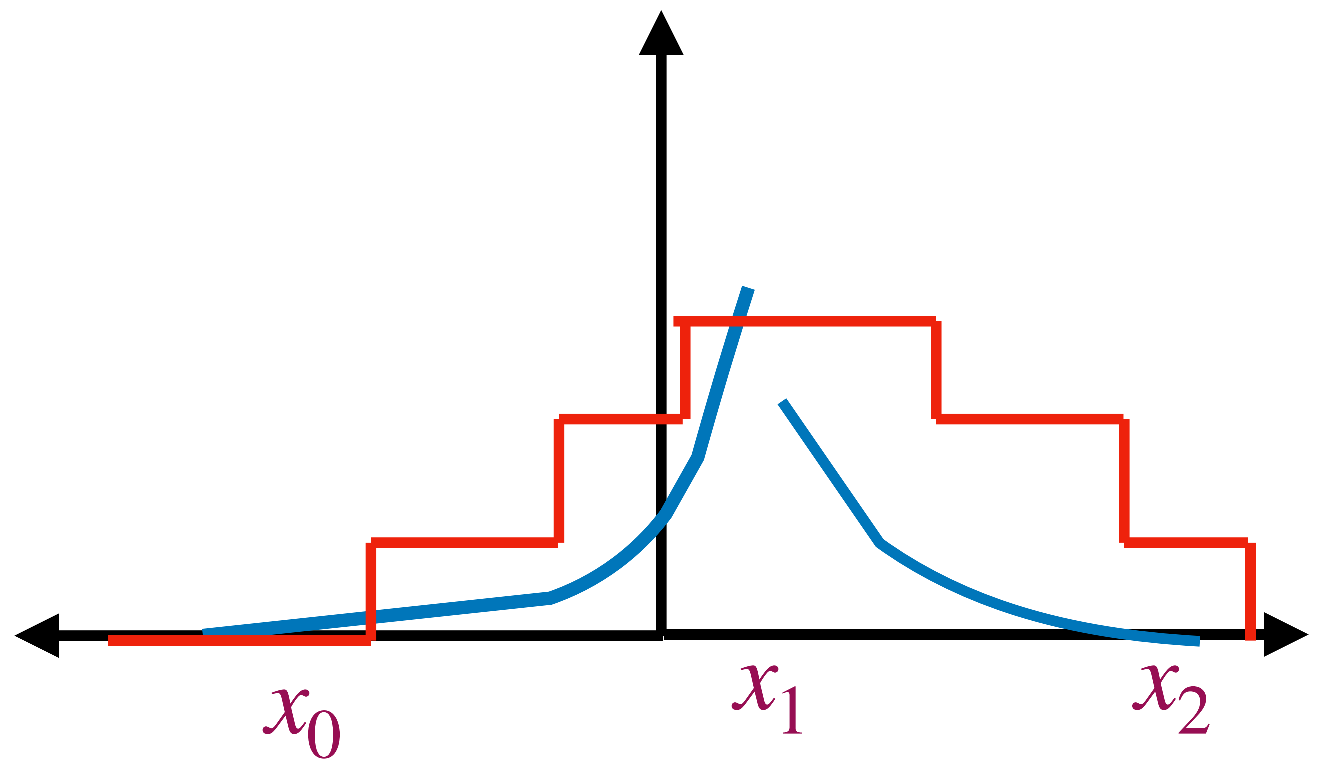
$$P_i^\star = (1-\alpha_i)\,\overline{P} + \alpha_i R_i\ ,\ \overline{P} \in \overline{\mathscr{O}}(D)$$

e.g. outliers $\exists\, S \subset \{1, \ldots, n\}\ :\ \forall i \in S, \alpha_i = 1$ and $\forall i \notin S, \alpha_i = 0$

$$\textbf{(AR)}\quad \mathbf{E}[d(\overline{P}, \widehat{P})] \leqslant \frac{2|S|}{n} + 48\sqrt{\frac{D}{n}} + \frac{\varepsilon}{n}$$

[ideal] Estimation of unimodal densities $\overline{M}/\overline{\mathscr{M}}$

$\overline{O}^0_D/\overline{\mathscr{O}}^0_D$ set of unimodal densities/distributions
constant on $\leqslant D$ bounded intervals



Proposition

> The elements of $\overline{\mathscr{O}}^0_D$ are extremal in $\overline{M}$ with degree $\leqslant 3(D+5)$

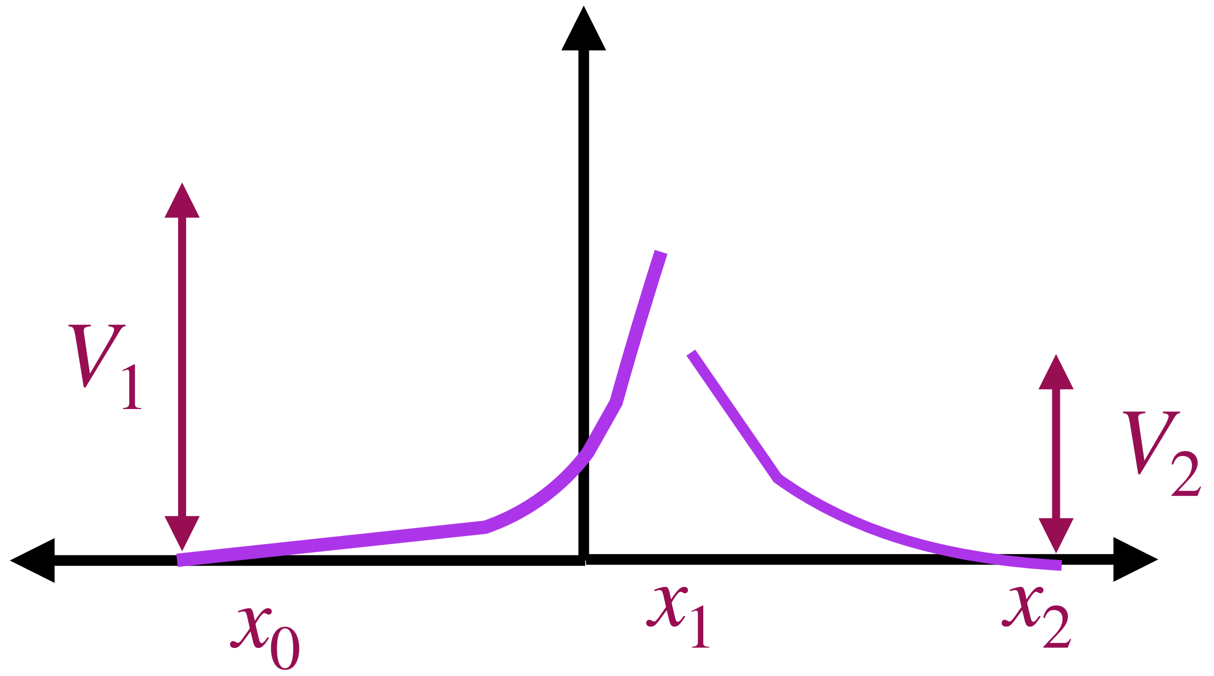
Theorem

> For all TV-estimator $\widehat{P}$,
> $$\mathbf{E}[d(\overline{P}^\star, \widehat{P})] \leqslant \inf_{D \geqslant 1} \left[3 \inf_{P \in \overline{\mathscr{O}}(D)} d(\overline{P}^\star, P) + 48\sqrt{3}\sqrt{\frac{D+5}{n}} \right] + \frac{\varepsilon}{n}$$

Eg 1. Bounded and compactly supported unimodal densities $\overline{M}$

(1) $f = w_1 p_1 \mathbf{1}_{(x_0,x_1)} + w_2 p_2 \mathbf{1}_{(x_1,x_2)}$ a.e. where $|x_i - x_{i-1}| = L_i$

$$V_i = \sup_{(x_{i-1},x_i)} p_i - \inf_{(x_{i-1},x_i)} p_i < \infty$$



$w_1 + w_2 = 1$

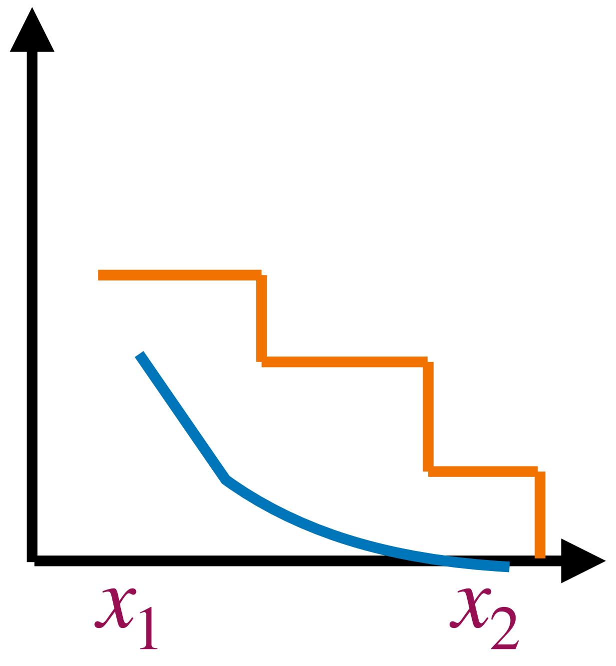
$$\overline{M}(R) \subset \overline{M} \qquad \inf_{f \text{ as } (1)} \left[\sqrt{w_1 \log(1 + L_1 V_1)} + \sqrt{w_2 \log(1 + L_2 V_2)}\right]^2 < R$$

Theorem

For all TV-estimator $\widehat{P}$,

$$\mathbf{E}[d(P^\star, \widehat{P})] \leqslant 41.3 \left(\frac{R}{n}\right)^{1/3} + 166.4 \sqrt{\frac{2}{n}} + \frac{\varepsilon}{n}$$

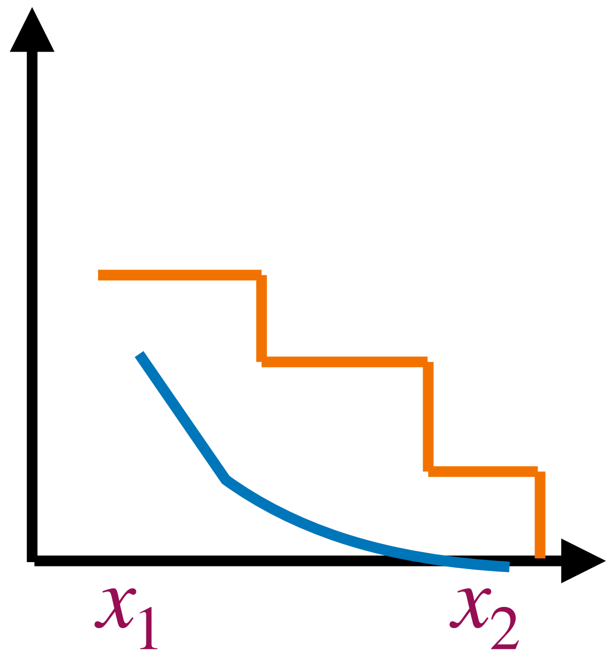
Particular case. Decreasing densities on an interval



(1) $f = p_2 \mathbf{1}_{(x_1,x_2)}$ a.e. $\quad \inf_{f \text{ as } (1)} \log(1 + LV) < R$

$$\mathbf{E}[d(P^\star, \widehat{P})] \leqslant 41.3 \left(\frac{R}{n}\right)^{1/3} + 166.4\sqrt{\frac{2}{n}} + \frac{\varepsilon}{n}$$

Particular case. Decreasing densities on an interval



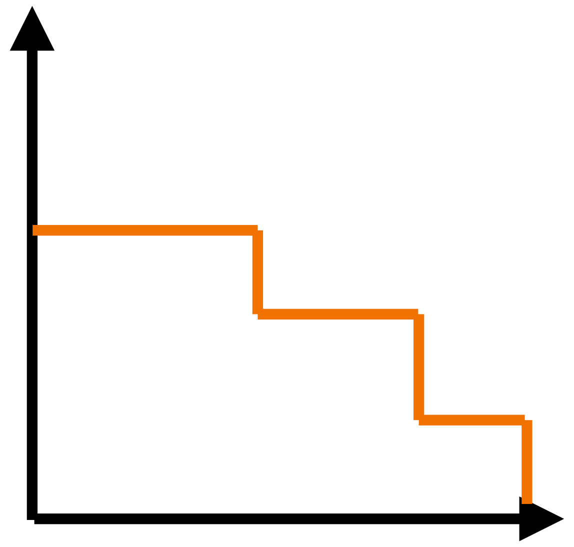
(1) $f = p_2 \mathbf{1}_{(x_1,x_2)}$ a.e. $\qquad \inf_{f \text{ as } (1)} \log(1+LV) < R$

$$\mathbf{E}[d(P^\star, \widehat{P})] \leqslant 41.3\underbrace{\cancel{\left(\frac{R}{n}\right)^{1/3}}}_{=0} + 166.4\sqrt{\frac{2}{n}} + \frac{\varepsilon}{n}$$

MLE

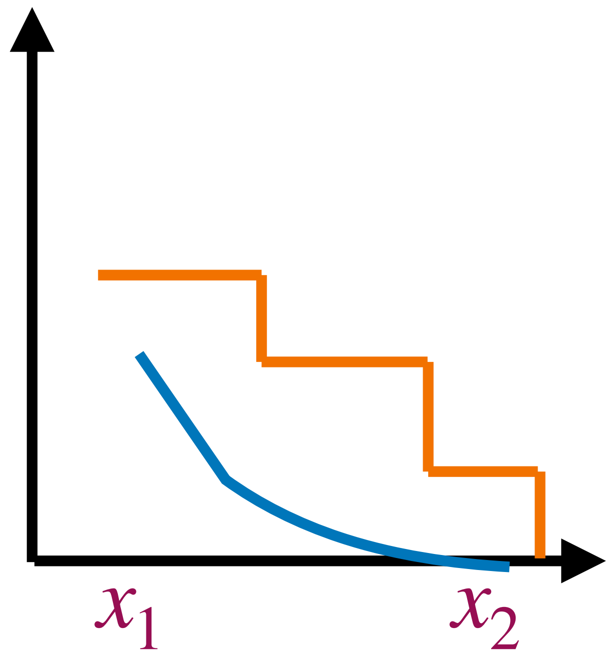
[Adapt] If $\overline{P}^\star = P^\star$ in $\overline{\mathscr{O}}(D)$, rate $\sim 1/\sqrt{n}$

$\forall i \in \{1,\ldots,D\}, V_i = 0 \Rightarrow R = 0$



Particular case. Decreasing densities on an interval

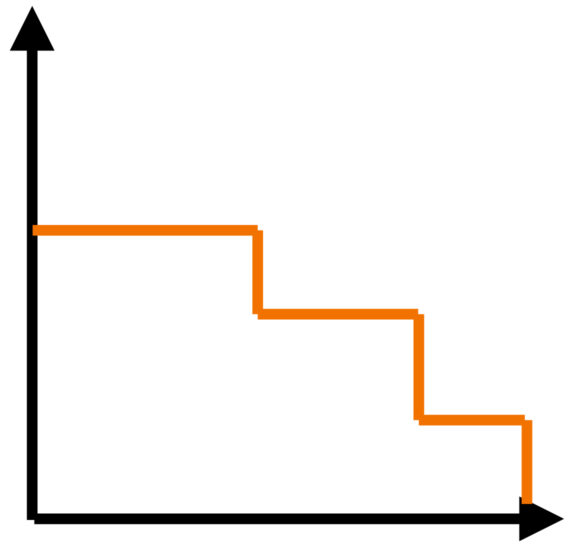
(1) $f = p_2 \mathbf{1}_{(x_1,x_2)}$ a.e. $\qquad \inf_{f \text{ as } (1)} \log(1+LV) < R$



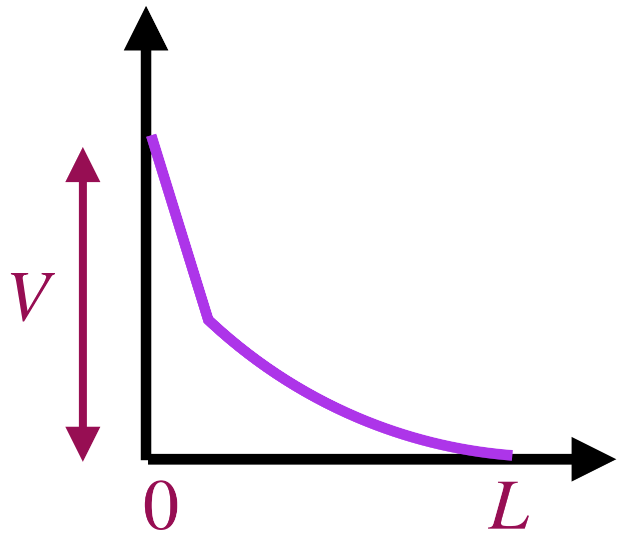
$$\mathbf{E}[d(P^\star, \widehat{P})] \leqslant 41.3\left(\frac{R}{n}\right)^{1/3} + 166.4\sqrt{\frac{2}{n}} + \frac{\varepsilon}{n}$$

MLE $\quad$ **[Adapt]** If $\overline{P}^\star = P^\star$ in $\overline{\mathcal{O}}(D)$, rate $\sim 1/\sqrt{n}$

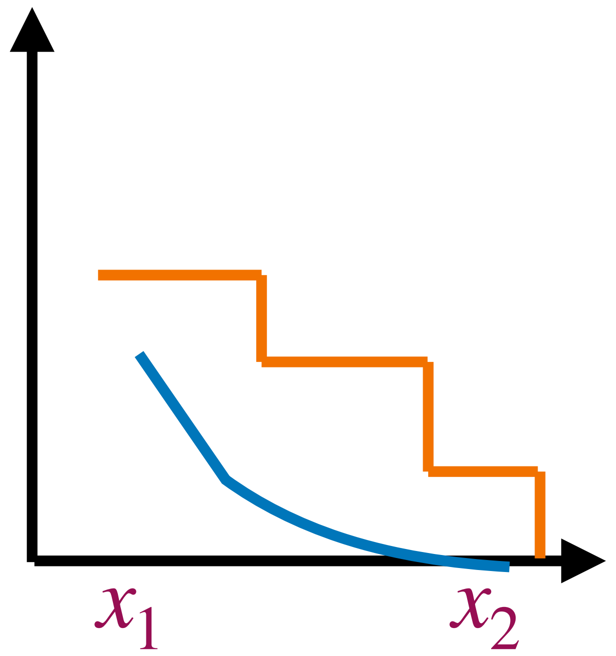
$\forall i \in \{1,\ldots,D\},\ V_i = 0 \Rightarrow R = 0$



MLE $\quad$ **[Opt]** rate $\sim [\log(1+LV)/n]^{1/3}$



Particular case. Decreasing densities on an interval



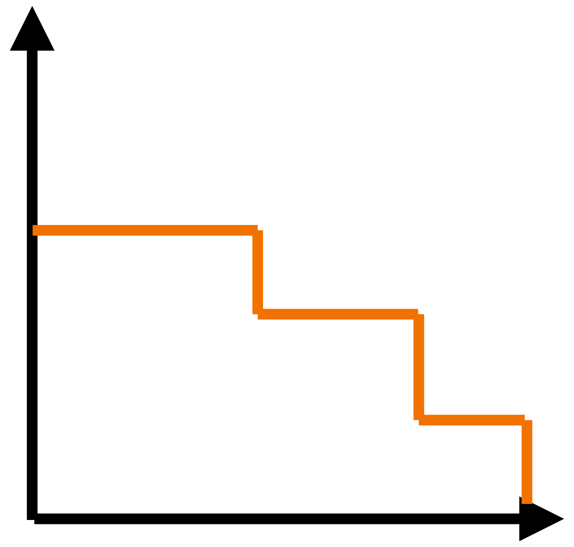
(1) $f = p_2 \mathbf{1}_{(x_1,x_2)}$ a.e. $\qquad \inf_{f \text{ as } (1)} \log(1 + LV) < R$

$$\mathbf{E}[d(P^\star, \widehat{P})] \leqslant 41.3 \left(\frac{R}{n}\right)^{1/3} + 166.4 \sqrt{\frac{2}{n}} + \frac{\varepsilon}{n}$$

✓ MLE

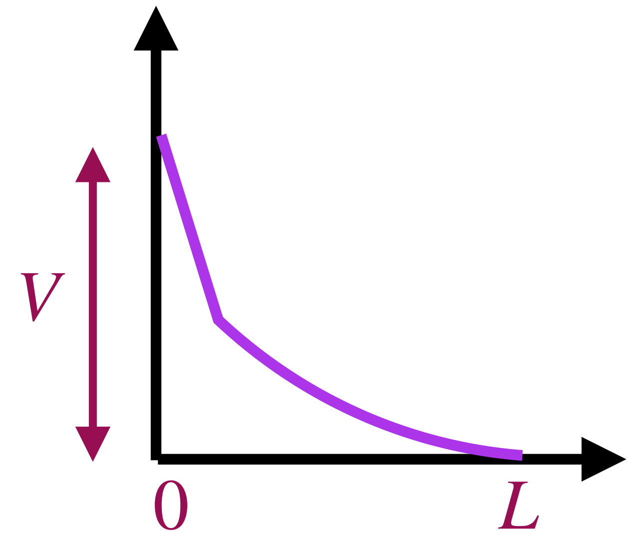
✓ **[Adapt]** If $\overline{P}^\star = P^\star$ in $\overline{\mathscr{O}}(D)$, rate $\sim 1/\sqrt{n}$

$\forall i \in \{1, \ldots, D\},\ V_i = 0 \Rightarrow R = 0$



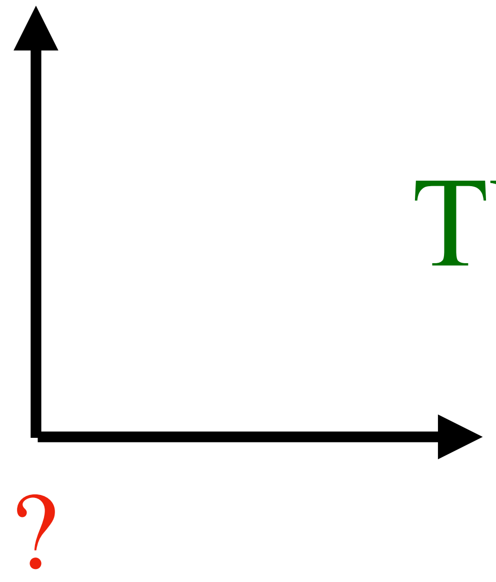
✓ MLE

✓ **[Opt]** rate $\sim [\log(1 + LV)/n]^{1/3}$



✗ MLE

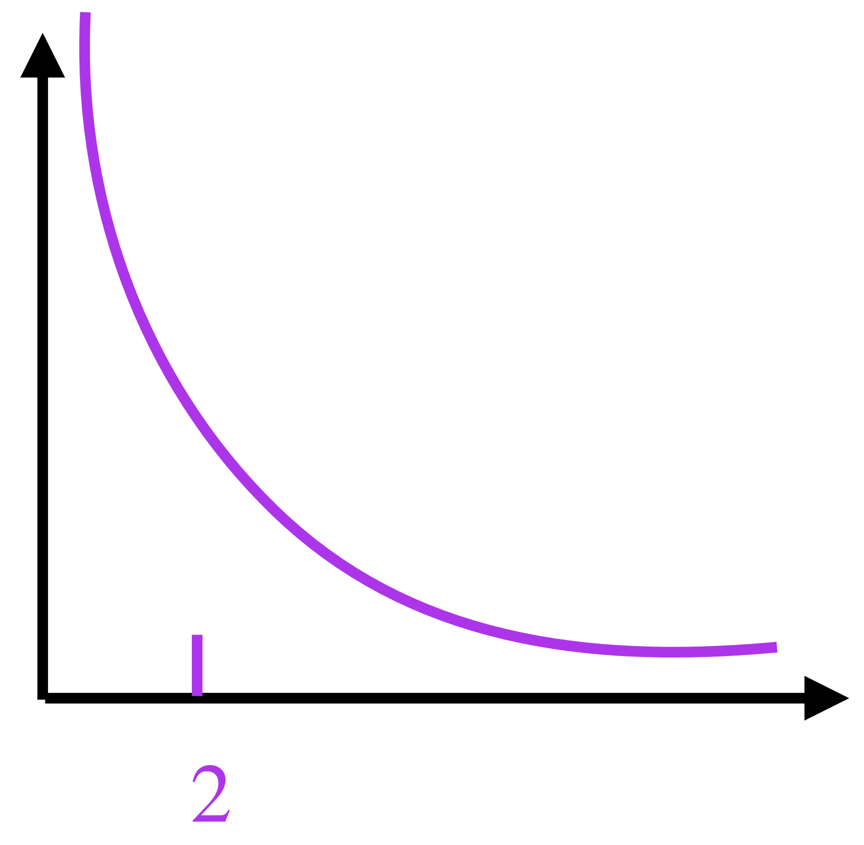
✓ x_1 can be unknown



TV-est. exists

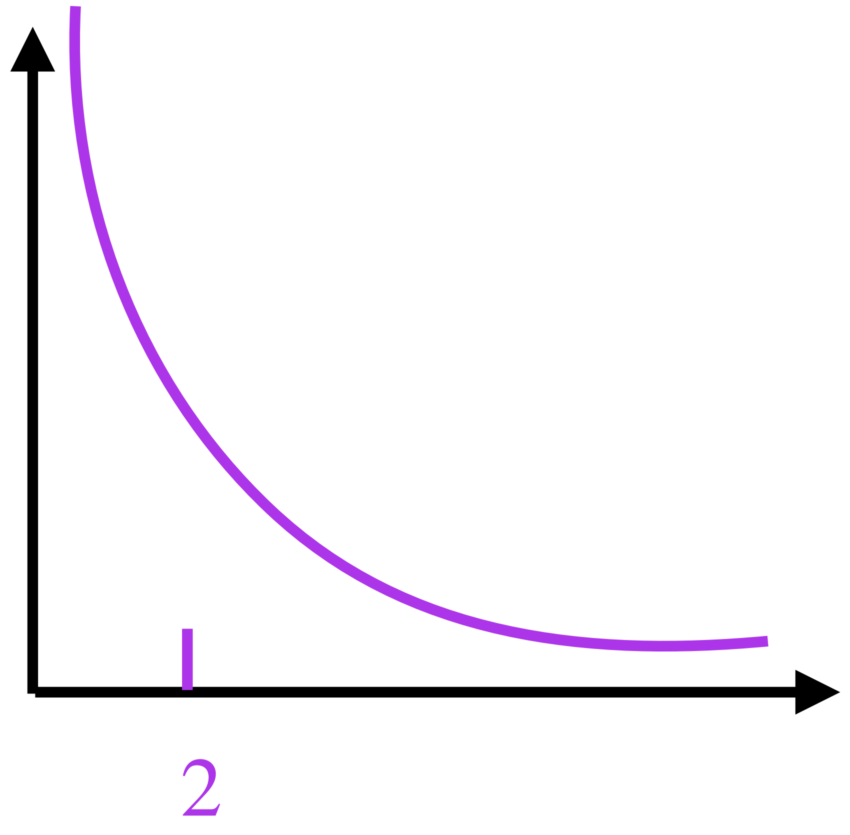
Eg 2. A non-compactly supported and unbounded monotone density

$$f(x) = \frac{c}{\sqrt{x}}\mathbf{1}_{(0,2)}(x) + \frac{d}{x\log(x)^2}\mathbf{1}_{[2,\infty)}(x)$$



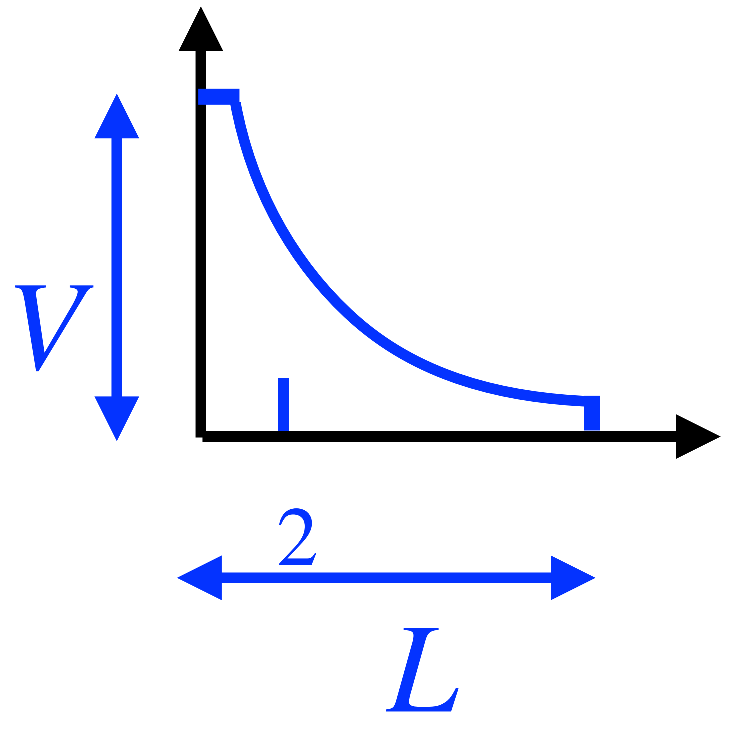
Eg 2. A non-compactly supported and unbounded monotone density

$$f(x) = \frac{c}{\sqrt{x}}\mathbf{1}_{(0,2)}(x) + \frac{d}{x\log(x)^2}\mathbf{1}_{[2,\infty)}(x)$$



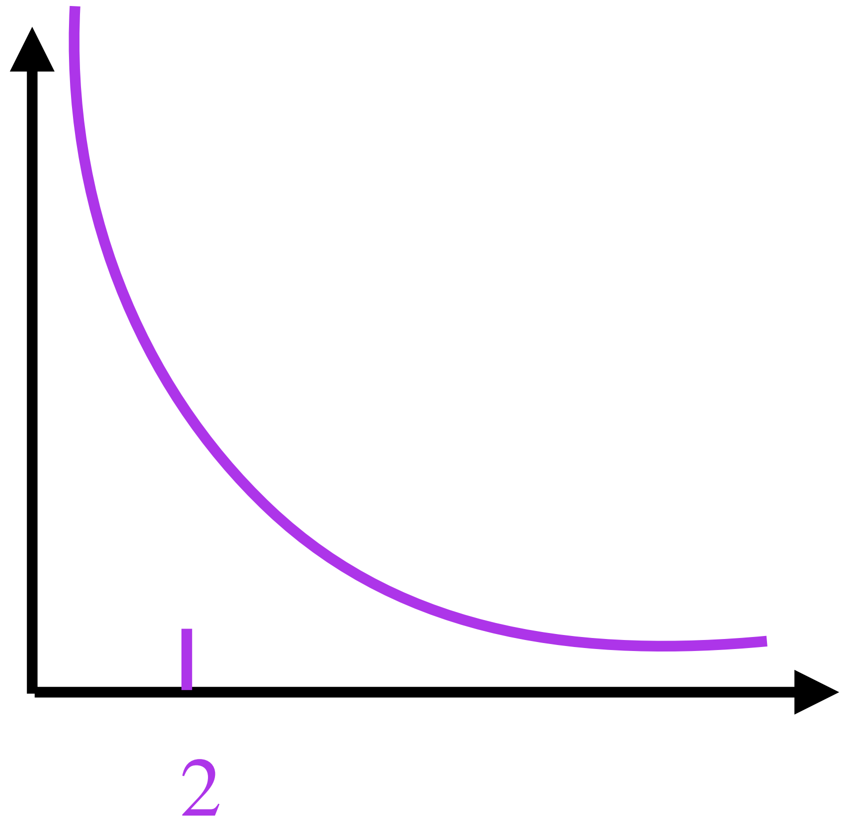
How to approximate f by an element of $\overline{M}(R)$ for some R ?

$$(1) \quad p = p_2\mathbf{1}_{(x_1,x_2)}$$



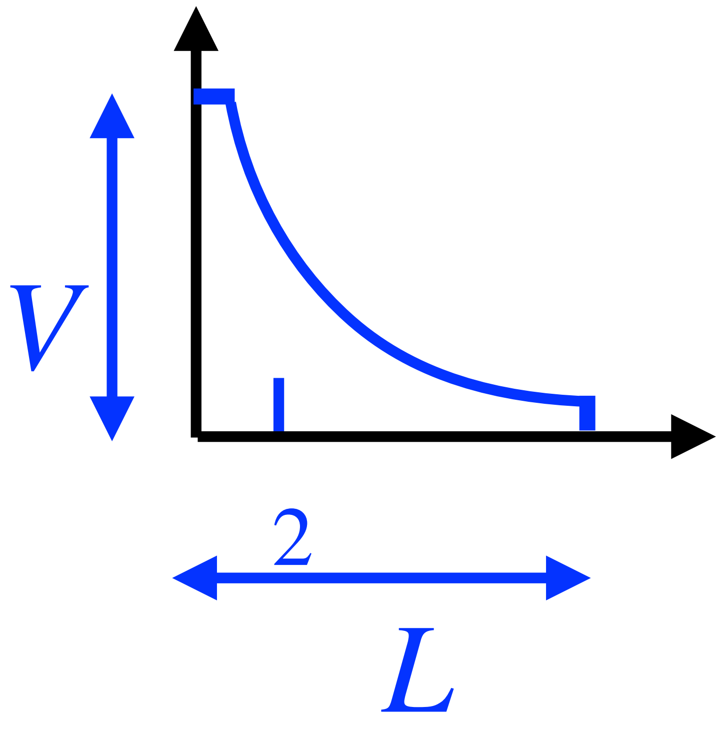
$$\log(1+LV) < R$$

Eg 2. A non-compactly supported and unbounded monotone density



$$f(x) = \frac{c}{\sqrt{x}}\mathbf{1}_{(0,2)}(x) + \frac{d}{x\log(x)^2}\mathbf{1}_{[2,\infty)}(x)$$

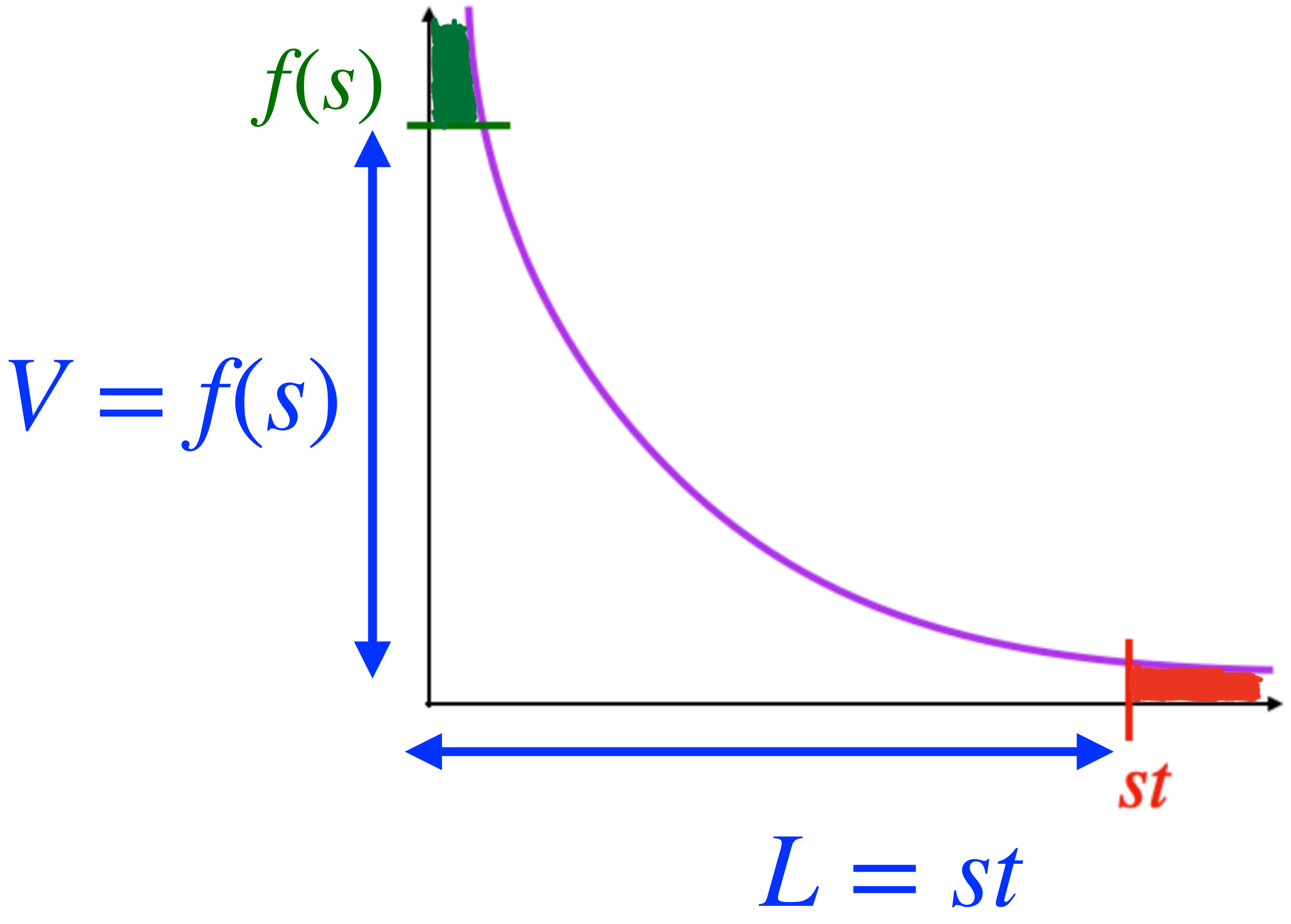
How to approximate f by an element of $\overline{M}(R)$ for some R ?



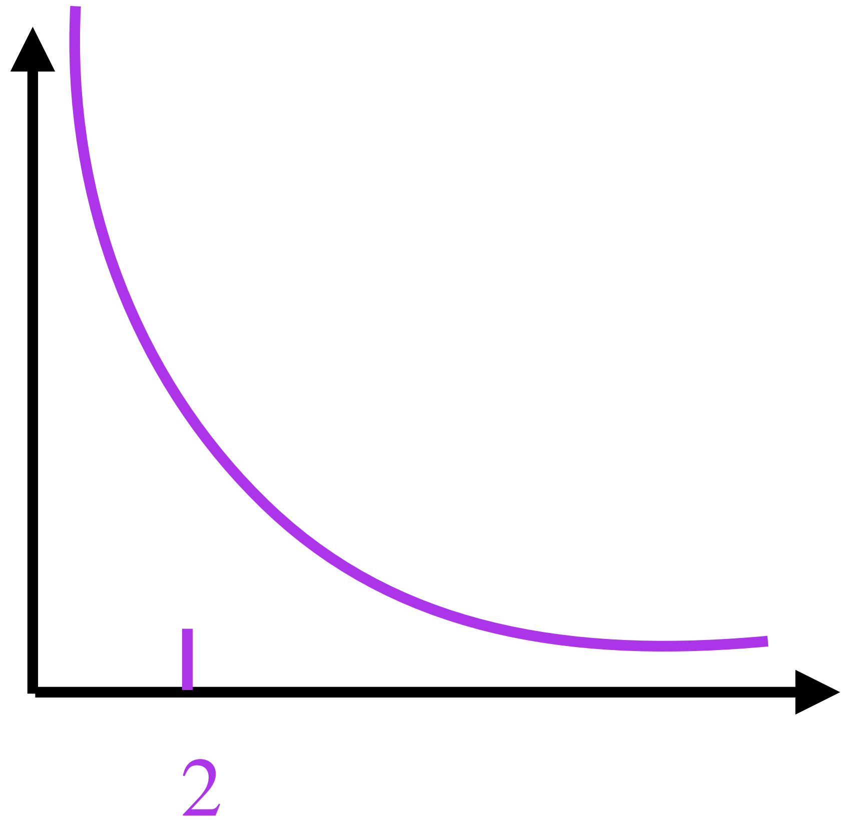
$$(1)\quad p = p_2\mathbf{1}_{(x_1,x_2)}$$

$\log(1+LV) < R$

Tail function $\quad \tau(f,t) = \inf_{s>0}\left[\ \tau_x(f,st) + \tau_y(f,f(s))\ \right]$

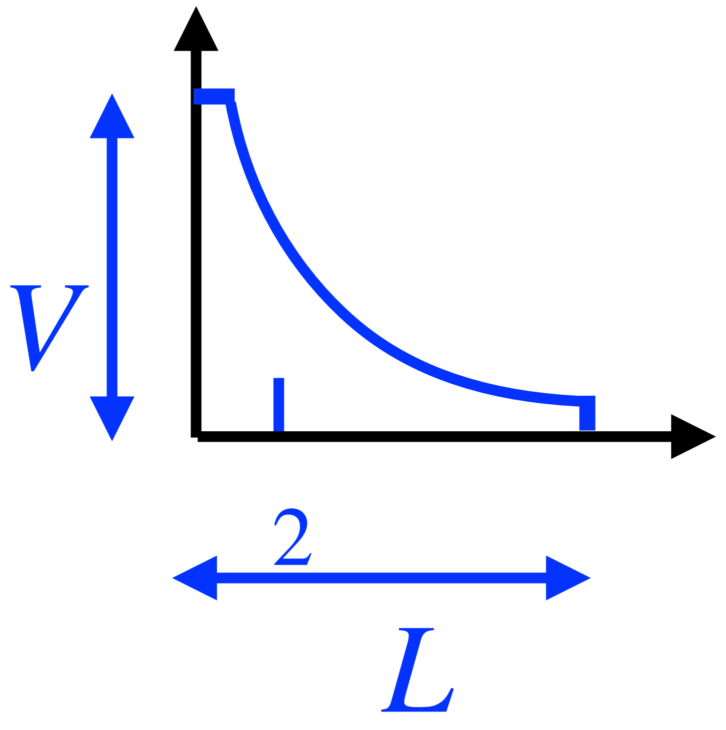


Eg 2. A non-compactly supported and unbounded monotone density



$$f(x) = \frac{c}{\sqrt{x}}\mathbf{1}_{(0,2)}(x) + \frac{d}{x\log(x)^2}\mathbf{1}_{[2,\infty)}(x)$$

How to approximate f by an element of $\overline{M}(R)$ for some R ?



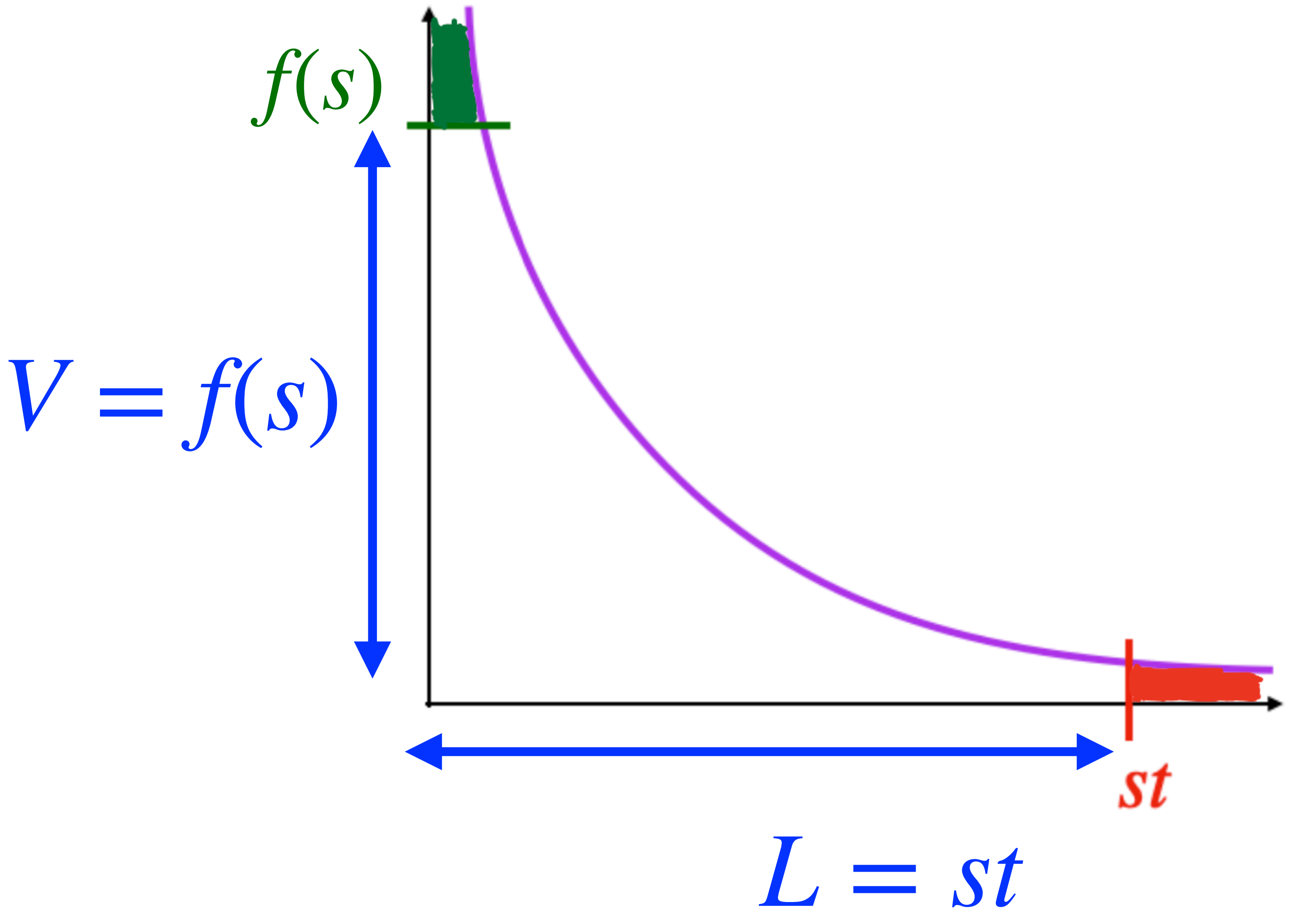
$$(1)\quad p = p_2\mathbf{1}_{(x_1,x_2)}$$

$\log(1+LV) < R$

Tail function $\tau(f,t) = \inf_{s>0}\left[\ \tau_x(f,st) + \tau_y(f,f(s))\ \right]$

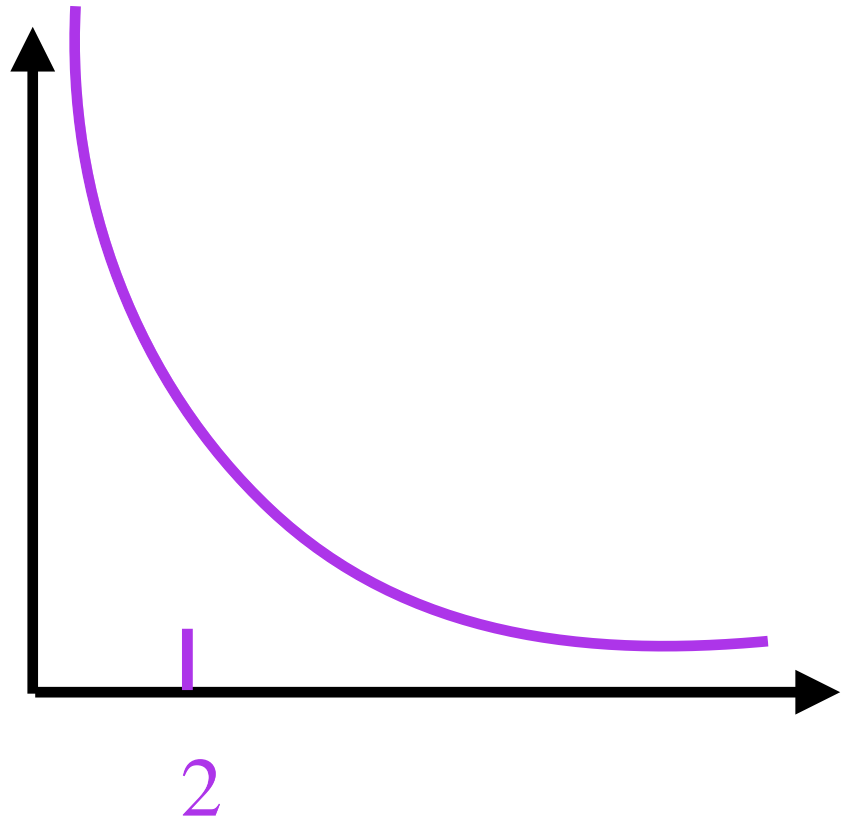
$t = \exp(R/2) - 1$

s.t. $\log(1 + f(s) \times st) \leqslant R$



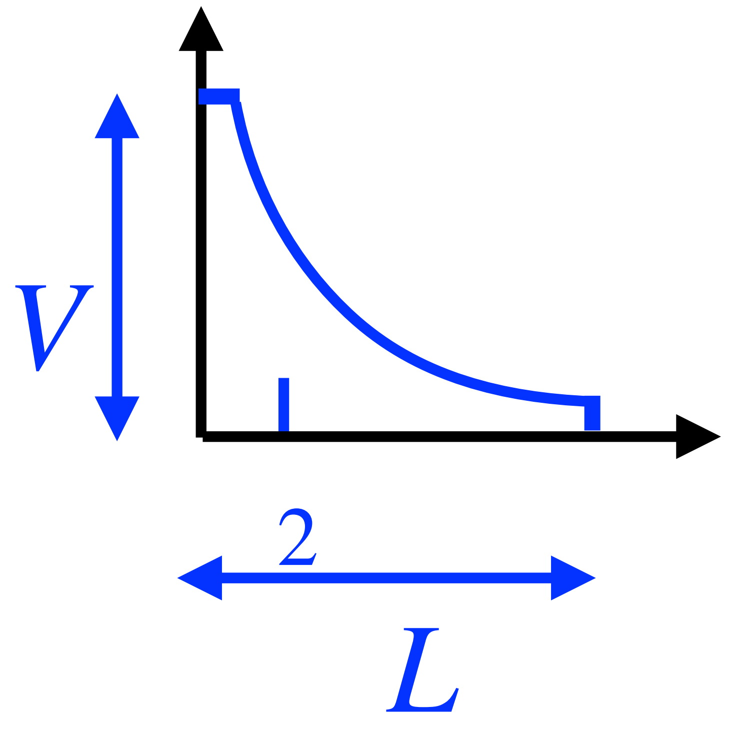
Eg 2. A non-compactly supported and unbounded monotone density

$$f(x) = \frac{c}{\sqrt{x}}\mathbf{1}_{(0,2)}(x) + \frac{d}{x\log(x)^2}\mathbf{1}_{[2,\infty)}(x)$$



How to approximate f by an element of $\overline{M}(R)$ for some R ?

$$(1) \quad p = p_2 \mathbf{1}_{(x_1,x_2)}$$

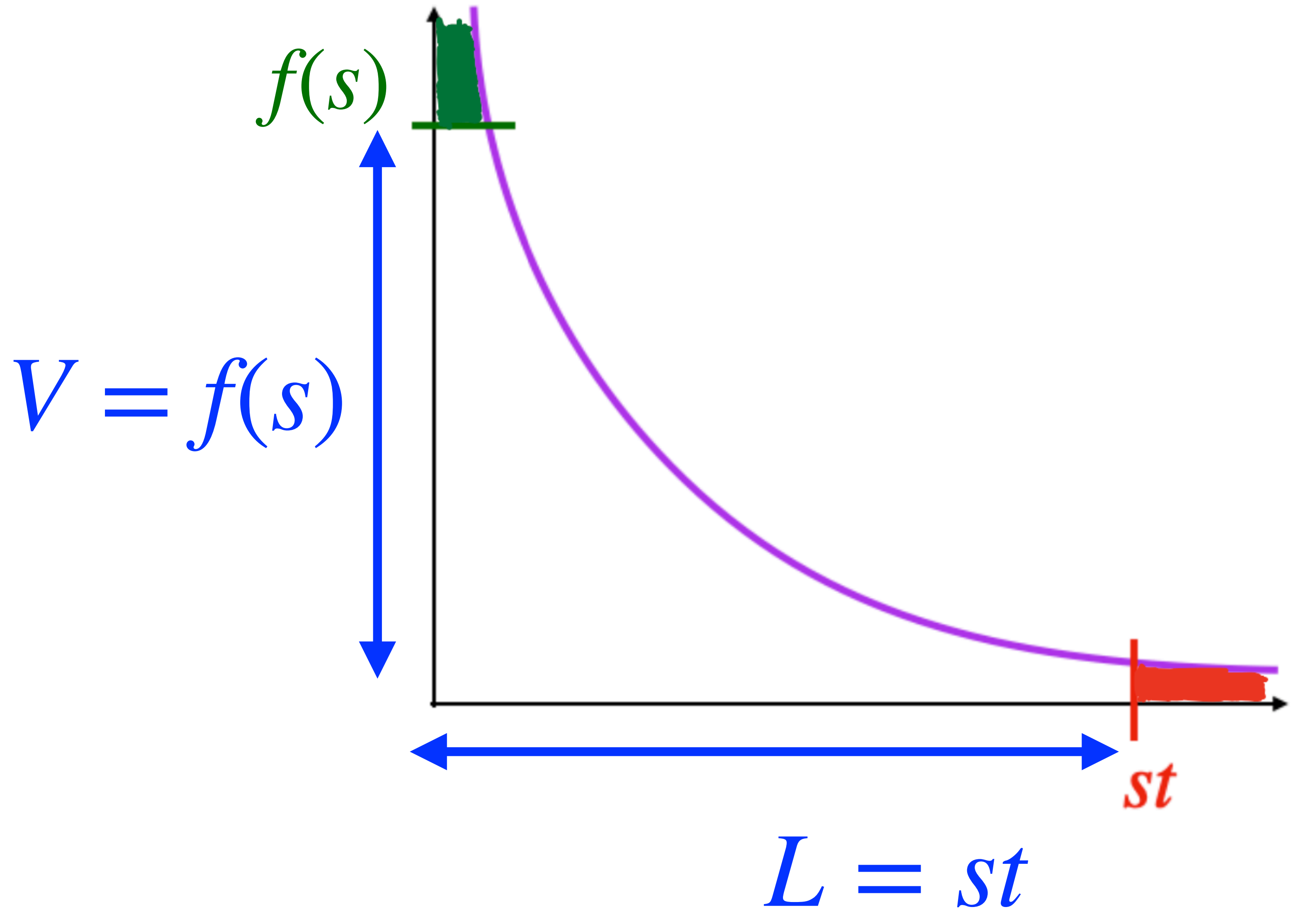


$\log(1+LV) < R$

Tail function $\quad \tau(f,t) = \inf_{s>0}\left[\ \tau_x(f,st) + \tau_y(f,f(s))\ \right]$

$t = \exp(R/2) - 1$

s.t. $\log(1 + f(s) \times st) \leqslant R$



Take $R = \kappa n^{1/4}$

For all TV-estimator $\widehat{P}$,

$$\mathbf{E}[d(P,\widehat{P})] \leqslant C\left[\frac{1}{n^{1/4}} + \sqrt{\frac{8}{n}}\right] + \frac{\varepsilon}{n}$$

Convex-concave densities

Model $\overline{M} =$ convex/concave + monotone densities on some bounded interval

Extremal points $\overline{\mathcal{O}}^1_D =$ piecewise linear elements of with D bounded affine pieces

Proposition

> The elements of $\overline{\mathcal{O}}^1_D$ are extremal in $\overline{M}$ with degree $\leqslant 2(D+5)$

Approx. error of a convex-concave monotone pdf

[Ass] Assume that p is a density s.t.

- p is supported on $[a, b]$ where $b - a = L > 0$
- $\sup_{(a,b)} p - \inf_{(a,b)} p = V < +\infty$
- Let also $\Gamma = \Gamma(p)$ be the linear index

$$\Gamma = 1 - \frac{1}{2}\left(\frac{|p_r'(a)| \wedge |p_l'(b)|}{\Delta} + \frac{\Delta}{|p_r'(a)| \vee |p_l'(b)|}\right)$$

where $\Delta = \dfrac{|p_l(b) - p_r(a)|}{b - a}$

Approximation error : theorem

Theorem

For any density p that satisfies **[Ass]**, there exists $\overline{p} \in \overline{\mathcal{O}}_D^1$ such that

$$d(P, \overline{P}) \leqslant \left[\left(1 + \sqrt{2\Gamma L V}\right)^{1/D} - 1\right]^2 \wedge 1$$

$$\leqslant 5.14 \frac{\log^2\left(1 + \sqrt{2\Gamma L V}\right)}{D^2}.$$

Global rates of convergence

Theorem

Let a density p that satisfies **[Ass]**. For all TV-estimator $\widehat{P}$,

(AR)
$$d(\widehat{P}, P) \leqslant 2d(P, \overline{P}^\star) + Cn^{-2/5}\log^{2/5}\left(1 + \sqrt{2\Gamma LV}\right)$$

$$d(\widehat{P}, \overline{P}^\star) \leqslant 3d(P, \overline{P}^\star) + Cn^{-2/5}\log^{2/5}\left(1 + \sqrt{2\Gamma LV}\right)$$

✔ Robust ✔ $n^{-2/5}$ rate ✔ log-dep on LV

Comparison to bibliography

Gao and Wellner (2009)

❌ Robust (MLE) ✅ Rate $n^{-2/5}$ ✅ log dep on LV (likely)

Chan et al (2014)

✅ Robust ✅ Rate $n^{-2/5}$ ❌ Concave only

Log-concave density estimation

Model $\overline{M}^{LC}$ = log-concave densities on $\mathbb{R}$,

$$p = \exp \phi,$$

$\phi : J \to \mathbb{R}$ continuous + concave, J open

Extremal points $\overline{\mathcal{O}}_D^{LC}$ = elements p of $\overline{M}^{LC}$ s.t. $\log p$ is piecewise affine or $= -\infty$, $D+1$pieces

Proposition

> The elements of $\overline{\mathcal{O}}_D^{LC}$ are extremal in $\overline{M}$ with degree $\leqslant 2(D+2)$

Global rates of convergence

Proposition

For any $p \in \overline{M}^{LC}$,

$$\inf_{q \in \overline{\mathcal{O}}^{LC}_{6D}} d(P, Q) \leqslant \frac{1}{D^2} \, .$$

It follows that

Corollary

If $\overline{P}^{\star} \in \overline{\mathscr{M}}^{LC}$, then

$$\mathbb{E}[d(\overline{P}^{\star}, \widehat{P})] \leqslant \frac{150}{n^{2/5}} + \frac{419}{\sqrt{n}} + \frac{\varepsilon}{n} \, .$$

Comparison to literature

Our estimator

✓ Robust ✓ Minimax rate ✓ Adaptive

MLE

✎ Robust? ✓ Minimax rate ✓ Adaptive

Chan et al (2014)

✓ Robust ✓ Minimax rate ✗ Adaptive

Conclusion

[1] TV-estimator VS MLE

- **[Adapt] [Opt]** as the MLE when it exists
- **[Vers]** Can be applied to density models for which the MLE does not exist

[2] Robustness

i.i.d. OK $P^\star \notin \overline{\mathscr{M}}$

- Misspecification
- Contamination

***Non*-i.i.d.**

- Equidistribution
- Outliers

[3] Other shape constraints : k-piecewise monotone densities, k-piecewise convex/concave densities

Some references

- Y. Baraud, *Tests and estimation strategies associated to some loss functions*, Proba. Th. and Rel. Fields, 2021.

- L. Birgé, *The Grenander estimator: a nonasymptotic approach*, Annals of Stat., 1989.

- S. Chan, I. Diakonikolas, R. Servedio, , *Efficient density estimation via polynomial approximation*, Proceedings og the 46th annual ACM symposium on Theory of computing

- L. Devroye and G. Lugosi, *Combinatorial methods in density estimation*, Springer Series in Stat 2001.

- F. Gao, J. A. Wellner, *On the rate of convergence of the MLE of a k-monotone density*. Sci. In China Sr. A, 2009.

Thanks for listening !