Estimation in the convolution structure density model.

Oleg V. Lepski
(Joint work with Thomas Willer)

Institut de Mathématiques de Marseille
Aix-Marseille Université

CIRM, July 11, 2017

Convolution structure density model.

- Observation $Z^{(n)} = (Z_1, \dots, Z_n)$, $Z_i \in \mathbb{R}^d, i = 1, \dots, n$, are **i.i.d.** random vectors with common density $\mathfrak{p}$ satisfying the following structural assumption

$$\mathfrak{p} = (1-\alpha)f + \alpha[f \star g], \quad f \in \mathbb{F}_g(R), \quad \alpha \in [0,1].$$

- $g \in \mathbb{L}_1(\mathbb{R}^d)$ and $\alpha \in [0,1]$ are **known**;
- $f \in \mathbb{F}_g(R)$, $R > 1$ **to be estimated**;

$$\mathbb{F}_g(R) = \Big\{ f \in \mathbb{B}_{1,d}(R) : (1-\alpha)f + \alpha[f \star g] \in \mathfrak{P}\big(\mathbb{R}^d\big) \Big\}$$

  - $\mathbb{B}_{1,d}(R)$ denotes the open ball of the radius **R** in $\mathbb{L}_1(\mathbb{R}^d)$;
  - $\mathfrak{P}\big(\mathbb{R}^d\big)$ is the set of all probability densities on $\mathbb{R}^d$;
  - $\big[f \star g\big](\cdot) = \int_{\mathbb{R}^d} f(\cdot - y) g(y) \mathrm{d}y$

Particular case: $f, g \in \mathfrak{P}(\mathbb{R}^d)$

- Observation $Z^{(n)} = (Z_1, \ldots, Z_n)$

$$Z_i = X_i + \varepsilon_i Y_i, \quad i = 1, \ldots, n$$

- $X_i \in \mathbb{R}^d, i = 1, \ldots, n$ are **i.i.d.** random vectors with common density **f to be estimated**;
- The noise variables $Y_i \in \mathbb{R}^d, i = 1, \ldots, n$, are **i.i.d.** random vectors with **known** common density g;
- $\varepsilon_i \in \{0, 1\}, i = 1, \ldots, n$, are **i.i.d.** Bernoulli random variables with $\mathbb{P}(\varepsilon_1 = 1) = \alpha$, $\alpha \in [0, 1]$ **is supposed to be known**;
- The sequences $\{X_i, i = 1, \ldots, n\}, \{Y_i, i = 1, \ldots, n\}$ and $\{\varepsilon_i, i = 1, \ldots, n\}$ are supposed to be mutually independent.

- $\alpha = 0$, **direct observations** $Z_i = X_i$;
- $\alpha = 1$, **density deconvolution** $Z_i = X_i + Y_i$;
- $\alpha \in (0, 1)$, **partially contaminated observations,**[Hesse].

Particular case: $f, g \in \mathfrak{P}(\mathbb{R}^d)$

- Observation $Z^{(n)} = (Z_1, \ldots, Z_n)$

$$Z_i = X_i + \varepsilon_i Y_i, \quad i = 1, \ldots, n$$

- $X_i \in \mathbb{R}^d, i = 1, \ldots, n$ are **i.i.d.** random vectors with common density **f to be estimated**;
- The noise variables $Y_i \in \mathbb{R}^d, i = 1, \ldots, n$, are **i.i.d.** random vectors with **known** common density **g**;
- $\varepsilon_i \in \{0, 1\}, i = 1, \ldots, n$, are **i.i.d.** Bernoulli random variables with $\mathbb{P}(\varepsilon_1 = 1) = \alpha$, $\alpha \in [0, 1]$ **is supposed to be known**;
- The sequences $\{X_i, i = 1, \ldots, n\}, \{Y_i, i = 1, \ldots, n\}$ and $\{\varepsilon_i, i = 1, \ldots, n\}$ are supposed to be mutually independent.

- $\alpha = 0$, **direct observations** $Z_i = X_i$;
- $\alpha = 1$, **density deconvolution** $Z_i = X_i + Y_i$;
- $\alpha \in (0, 1)$, **partially contaminated observations,** **[Hesse].**

- Convolution structure density model

$$\mathfrak{p} = (1-\alpha)f + \alpha[f \star g], \quad f \in \mathbb{F}_g(R), \ \ \alpha \in [0,1].$$

We want to estimate **f** using observations $Z^{(n)} = (Z_1, \dots, Z_n)$.

- Estimator is $Z^{(n)}$-measurable map $\hat{f} : (\mathbb{R}^d)^n \to \mathbb{L}_p\big(\mathbb{R}^d\big)$.
- Accuracy of an estimator $\hat{f}$ is measured by the $\mathbb{L}_p$–risk

$$\mathcal{R}_n^{(p)}[\hat{f}, f] := \Big(\mathbb{E}_f \|\hat{f} - f\|_p^q\Big)^{1/q}, \ \ q \geq 1$$

- $\mathbb{E}_f$ denotes expectation with respect to the probability measure $\mathbb{P}_f$ of the observations $Z^{(n)} = (Z_1, \dots, Z_n)$.
- $\|\lambda\|_p^p = \int_{\mathbb{R}^d} |\lambda|^p \nu_d(\mathrm{d}x), \ \ 1 \leq p < \infty;$
- $\|\lambda\|_\infty = \sup_{x \in \mathbb{R}^d} |\lambda(x)|.$

Assumptions on the function g.

- Later on $\check{Q}$ denotes the Fourier transform of $\mathbf{Q}$.

Assumption 1. $\forall t \in \mathbb{R}^d$

- If $\alpha \in [0,1)$ then there exists $\varepsilon > 0$ such that

$$\left|1-\alpha+\alpha\check{g}(t)\right| \geq \varepsilon.$$

- If $\alpha = 1$ then there exist $\Upsilon_0, \Upsilon_1 > 0$ and $\vec{\mu} \in (0,\infty)^d$ s.t.

$$\Upsilon_0 \prod_{j=1}^{d}(1+t_j^2)^{-\frac{\mu_j}{2}} \leq |\check{g}(t)| \leq \Upsilon_1 \prod_{j=1}^{d}(1+t_j^2)^{-\frac{\mu_j}{2}}.$$

- Assumption 1 is very week if $\alpha \in [0,1)$ and it holds with $\varepsilon = 1-\alpha$ if $\check{g}$ is a real positive function (in particular for centered multivariate Laplace and Gaussian laws).
- If g is a probability density then Assumption 1 **always** holds with $\varepsilon = 1-2\alpha$ if $\alpha < 1/2$.
- In the case $\alpha = 1$ this assumption is referred to **moderately ill-posed** statistical problem.

Family of kernel-based estimators.

- For any $\vec{h} \in \mathcal{H}^d$ let $M(\cdot, \vec{h})$ satisfy the operator equation

$$K_{\vec{h}}(\cdot) = (1-\alpha)M(\cdot, \vec{h}) + \alpha \int_{\mathbb{R}^d} g(t-\cdot)M(t, \vec{h})\mathrm{d}t$$

  - $\mathcal{H}^d$ is the diadic grid in $(0, \infty)^d$;
  - $K_{\vec{h}}(y) = \left[\prod_{j=1}^d h_j^{-1}\right] K(y_1/h_1, \ldots, y_d/h_d), \; y \in \mathbb{R}^d$
  - Kernel $K \in \mathbb{C}(\mathbb{R}^d) \cap \mathbb{L}_1(\mathbb{R}^d)$ is such that $\int_{\mathbb{R}} K = 1$ and

Assumption 2. $\exists \mathrm{k}_1, \mathrm{k}_2 > 0$ **such that**

1. $\int_{\mathbb{R}^d} |\check{K}(t)| \prod_{j=1}^d (1+t_j^2)^{\frac{\mu_j(\alpha)}{2}} \mathrm{d}t \le \mathrm{k}_1.$
2. $\int_{\mathbb{R}^d} |\check{K}(t)|^2 \prod_{j=1}^d (1+t_j^2)^{\mu_j(\alpha)} \mathrm{d}t \le \mathrm{k}_2^2.$

- $\vec{\mu}(\alpha) = \vec{\mu}$ if $\alpha = 1$ and $\vec{\mu}(\alpha) = 0$ if $\alpha \neq 1$.

Family of kernel-based estimators.

- For any $\vec{h} \in \mathcal{H}^d$ let $M(\cdot, \vec{h})$ satisfy the operator equation

$$K_{\vec{h}}(\cdot) = (1-\alpha)M(\cdot, \vec{h}) + \alpha \int_{\mathbb{R}^d} g(t-\cdot)M(t, \vec{h})\mathrm{d}t$$

  - $\mathcal{H}^d$ is the diadic grid in $(0, \infty)^d$;
  - $K_{\vec{h}}(y) = \left[\prod_{j=1}^d h_j^{-1}\right] K(y_1/h_1, \ldots, y_d/h_d), \; y \in \mathbb{R}^d$

- Kernel-based estimator

$$\widehat{f}_{\vec{\mathrm{h}}}(x) = n^{-1} \sum_{i=1}^n M(Z_i - x, \vec{\mathrm{h}})$$

Objective is to propose for any given $x \in \mathbb{R}^d$ a data-driven selection rule from the family of kernel-based estimators

$$\mathcal{F}(\mathcal{H}^d) = \{\widehat{f}_{\vec{\mathrm{h}}}(\cdot), \; \vec{\mathrm{h}} \in \mathbb{H}\}$$

  - $\mathbb{H}$ is an arbitrary subset of $\mathcal{H}^d$.

Pointwise selection rule.

- Set $\forall \vec{h} \in \mathbb{H}$, $\forall x \in \mathbb{R}^d$

$$\widehat{\mathcal{R}}_{\vec{h}}(x) = \sup_{\vec{\eta}\in\mathbb{H}} \Big[\big|\widehat{f}_{\vec{h}\vee\vec{\eta}}(x) - \widehat{f}_{\vec{\eta}}(x)\big| - 4\widehat{U}_n\big(x,\vec{h}\vee\vec{\eta}\big) - 4\widehat{U}_n\big(x,\vec{\eta}\big)\Big]_+;$$

$$\vec{\mathrm{h}}(x) = \arg\inf_{\vec{h}\in\mathbb{H}} \Big[\widehat{\mathcal{R}}_{\vec{h}}(x) + 8\widehat{U}^*_n\big(x,\vec{h}\big)\Big].$$

- Our final estimator is $\boxed{\widehat{f}_{\vec{\mathrm{h}}(x)}(x),\ \ x \in \mathbb{R}^d}$

- $\widehat{U}^*_n\big(x,\vec{h}\big) = \sup_{\vec{\eta}\in\mathbb{H}:\ \vec{\eta}\geq\vec{h}} \widehat{U}_n\big(x,\vec{\eta}\big);$

- $\widehat{U}_n\big(x,\vec{\mathrm{h}}\big) = \sqrt{\frac{2\lambda_n\big(\vec{\mathrm{h}}\big)\widehat{\sigma}^2\big(x,\vec{\mathrm{h}}\big)}{n}} + \frac{4M_\infty\lambda_n\big(\vec{\mathrm{h}}\big)}{3n\prod_{j=1}^d \mathrm{h}_j(\mathrm{h}_j\wedge 1)^{\mu_j(\alpha)}};$

$$\widehat{\sigma}^2\big(x,\vec{\mathrm{h}}\big) = \frac{1}{n}\sum_{i=1}^n M^2\big(Z_i - x,\vec{\mathrm{h}}\big);$$

$$M_\infty = \Big[(2\pi)^{-d}\big\{\varepsilon^{-1}\big\|\check{K}\big\|_1 1_{\alpha\neq 1} + \Upsilon_0^{-1}\mathrm{k}_1 1_{\alpha=1}\big\}\Big] \vee 1.$$

$$\lambda_n\big(\vec{\mathrm{h}}\big) = 4\ln(M_\infty) + 6\ln(n) + (8p+26)\sum_{j=1}^d \big[1+\mu_j(\alpha)\big]\big|\ln(\mathrm{h}_j)\big|$$

Pointwise selection rule.

- Set $\forall \vec{h} \in \mathbb{H}$, $\forall x \in \mathbb{R}^d$

$$\widehat{\mathcal{R}}_{\vec{h}}(x) = \sup_{\vec{\eta}\in\mathbb{H}} \Big[\big| \widehat{f}_{\vec{h}\vee\vec{\eta}}(x) - \widehat{f}_{\vec{\eta}}(x) \big| - 4\widehat{U}_n\big(x, \vec{h}\vee\vec{\eta}\big) - 4\widehat{U}_n\big(x,\vec{\eta}\big) \Big]_+;$$

$$\vec{\mathrm{h}}(x) = \arg\inf_{\vec{h}\in\mathbb{H}} \Big[\widehat{\mathcal{R}}_{\vec{h}}(x) + 8\widehat{U}_n^*\big(x,\vec{h}\big) \Big].$$

- Our final estimator is $\boxed{\widehat{f}_{\vec{\mathrm{h}}(x)}(x), \ \ x \in \mathbb{R}^d}$
- $\widehat{U}_n^*\big(x,\vec{h}\big) = \sup_{\vec{\eta}\in\mathbb{H}:\ \vec{\eta}\geq\vec{h}} \widehat{U}_n\big(x,\vec{\eta}\big);$
- $\widehat{U}_n\big(x,\vec{\mathrm{h}}\big) = \sqrt{\frac{2\lambda_n\big(\vec{\mathrm{h}}\big)\widehat{\sigma}^2\big(x,\vec{\mathrm{h}}\big)}{n}} + \frac{4M_\infty\lambda_n\big(\vec{\mathrm{h}}\big)}{3n\prod_{j=1}^d \mathrm{h}_j(\mathrm{h}_j\wedge 1)^{\mu_j(\alpha)}};$

$$\widehat{\sigma}^2\big(x,\vec{\mathrm{h}}\big) = \frac{1}{n}\sum_{i=1}^n M^2\big(Z_i - x, \vec{\mathrm{h}}\big);$$

$$M_\infty = \Big[(2\pi)^{-d}\big\{\varepsilon^{-1}\big\|\check{K}\big\|_1 1_{\alpha\neq 1} + \Upsilon_0^{-1}\mathrm{k}_1 1_{\alpha=1}\big\}\Big] \vee 1.$$

$$\lambda_n\big(\vec{\mathrm{h}}\big) = 4\ln(M_\infty) + 6\ln(n) + (8p+26)\sum_{j=1}^d \big[1+\mu_j(\alpha)\big]\big|\ln(\mathrm{h}_j)\big|$$

$\mathbb{L}_p$-norm oracle inequality

- $B^*_{\vec{h}}(x,f) = 2\sup_{\vec{\eta}\in\mathbb{H}} \left|B_{\vec{h}\vee\vec{\eta}}(x,f) - B_{\vec{\eta}}(x,f)\right| + B_{\vec{h}}(x,f)$;
  - $B_{\vec{h}}(x,f) = \left|\int_{\mathbb{R}^d} K_{\vec{h}}(t-x)f(t)\mathrm{d}t - f(x)\right|$;
- $U^*_n(x,\vec{h}) = \sup_{\vec{\eta}\in\mathcal{H}^d:\ \vec{\eta}\geq\vec{h}} U_n(x,\vec{\eta})$;
  - $U_n(x,\vec{h}) = \sqrt{\frac{2\lambda_n(\vec{h})\sigma^2(x,\vec{h})}{n}} + \frac{4M_\infty\lambda_n(\vec{h})}{3n\prod_{j=1}^d h_j(h_j\wedge 1)^{\mu_j(\alpha)}}$
  - $\sigma^2(x,\vec{h}) = \int_{\mathbb{R}^d} M^2(t-x,\vec{h})\mathfrak{p}(t)\nu_d(\mathrm{d}t)$;

Theorem 1. Let Assumptions 1 and 2 be fulfilled.

Then $\forall\mathbb{H}\subseteq\mathcal{H}^d$, $n\geq 3$, $p\in[1,\infty)$, $\forall f\in\mathbb{F}_g(\infty)$

$$\mathcal{R}^{(p)}_n\left[\widehat{f}_{\vec{h}(\cdot)}; f\right] \leq \left\| \inf_{\vec{h}\in\mathbb{H}} \left\{ B^*_{\vec{h}}(\cdot,f) + 49U^*_n(\cdot,\vec{h}) \right\} \right\|_p + C_p n^{-1}.$$

- C_p is independent of f, n and $\mathbb{H}$ (depend on K, g, p and d).

$\mathbb{L}_p$-norm oracle inequality

- $B^*_{\vec{h}}(x,f) = 2\sup_{\vec{\eta}\in\mathbb{H}} \left| B_{\vec{h}\vee\vec{\eta}}(x,f) - B_{\vec{\eta}}(x,f) \right| + B_{\vec{h}}(x,f)$;
  - $B_{\vec{h}}(x,f) = \left| \int_{\mathbb{R}^d} K_{\vec{\mathrm{h}}}(t-x)f(t)\mathrm{d}t - f(x) \right|$;
- $U^*_n\big(x,\vec{h}\big) = \sup_{\vec{\eta}\in\mathcal{H}^d:\ \vec{\eta}\geq\vec{h}} U_n\big(x,\vec{\eta}\big)$;
  - $U_n\big(x,\vec{h}\big) = \sqrt{\frac{2\lambda_n\left(\vec{h}\right)\sigma^2\left(x,\vec{h}\right)}{n}} + \frac{4M_\infty\lambda_n\left(\vec{h}\right)}{3n\prod_{j=1}^d h_j(h_j\wedge 1)^{\mu_j(\alpha)}}$
  - $\sigma^2\big(x,\vec{h}\big) = \int_{\mathbb{R}^d} M^2\big(t-x,\vec{h}\big)\mathfrak{p}(t)\nu_d(\mathrm{d}t)$;

Theorem 1. Let Assumptions 1 and 2 be fulfilled.

Then $\forall\mathbb{H}\subseteq\mathcal{H}^d, n\geq 3, p\in[1,\infty)$, $\forall f\in\mathbb{F}_g(\infty)$

$$\mathcal{R}_n^{(p)}\left[\widehat{f}_{\vec{\mathrm{h}}(\cdot)}; f\right] \leq \left\| \inf_{\vec{h}\in\mathbb{H}} \left\{ B^*_{\vec{h}}(\cdot,f) + 49U^*_n\big(\cdot,\vec{h}\big) \right\} \right\|_p + C_p n^{-1}.$$

- C_p is independent of f, n and $\mathbb{H}$ (depend on K, g, p and d).

Adaptive minimax approach.

Let $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$ be a collection of subsets of $\mathbb{F}_{g}(R)$.

- $\varphi_n(\mathbb{F}_\theta) = \inf\limits_{\hat{f}} \mathcal{R}_n^{(p)}[\hat{f}, \mathbb{F}_\theta]$ ($\mathbb{L}_p$–minimax risk)
  - $\mathcal{R}_n^{(p)}\left[\hat{f}, \mathbb{F}_\theta\right] := \sup\limits_{f\in\mathbb{F}_\theta} \mathcal{R}_n^{(p)}[\hat{f}, f]$
  - Infimum is taken over all possible estimators.
- $(\theta, R) \in \Theta \times (1, \infty)$–nuisance parameter.
- $\psi_n(\mathbb{F}_\theta) = \sup\limits_{f\in\mathbb{F}_\theta} \underbrace{\left\| \inf\limits_{\vec{h}\in\mathbb{H}} \left\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n(\cdot, \vec{h}) \right\} \right\|_p}_{\text{main term in the r.h.s. of the oracle inequality}} + C_p n^{-1}.$
- To study the asymptotics $\psi_n(\Sigma_\theta)\varphi_n^{-1}(\Sigma_\theta)$, $n \to \infty$

Since the construction of the estimator $\widehat{f}_{\vec{h}(\cdot)}$ is independent of (θ, R) and p we can analyze its adaptivity under an arbitrary $\mathbb{L}_p$-loss over an arbitrary scale of functional classes $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$. But $\sup_{f\in\mathbb{F}_\theta} \left\| \inf_{\vec{h}\in\mathbb{H}} \left\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n(\cdot, \vec{h}) \right\} \right\|_p$ is not easy to analyze!

Adaptive minimax approach.

Let $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$ be a collection of subsets of $\mathbb{F}_g(R)$.

- $\varphi_n(\mathbb{F}_\theta) = \inf_{\hat{f}} \mathcal{R}_n^{(p)}[\hat{f}, \mathbb{F}_\theta]$ ($\mathbb{L}_p$–minimax risk)
  - $\mathcal{R}_n^{(p)}\left[\hat{f}, \mathbb{F}_\theta\right] := \sup_{f\in\mathbb{F}_\theta} \mathcal{R}_n^{(p)}[\hat{f}, f]$
  - Infimum is taken over all possible estimators.
- $(\theta, R) \in \Theta \times (1, \infty)$–nuisance parameter.
- $$\psi_n(\mathbb{F}_\theta) = \sup_{f\in\mathbb{F}_\theta} \underbrace{\left\| \inf_{\vec{h}\in\mathbb{H}} \left\{\mathcal{B}_{\vec{h}}(\cdot, f) + U_n\big(\cdot, \vec{h}\big)\right\}\right\|_p}_{\text{main term in the r.h.s. of the oracle inequality}} + C_p n^{-1}.$$
- To study the asymptotics $\psi_n(\Sigma_\theta)\varphi_n^{-1}(\Sigma_\theta)$, $\mathbf{n} \to \infty$

Since the construction of the estimator $\widehat{f_{\vec{h}(\cdot)}}$ is independent of (θ, R) and p we can analyze its adaptivity under an arbitrary $\mathbb{L}_p$-loss over an arbitrary scale of functional classes $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$. But $\sup_{f\in\mathbb{F}_\theta} \left\| \inf_{\vec{h}\in\mathbb{H}} \left\{\mathcal{B}_{\vec{h}}(\cdot, f) + U_n\big(\cdot, \vec{h}\big)\right\}\right\|_p$ is not easy to analyze!

Adaptive minimax approach.

Let $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$ be a collection of subsets of $\mathbb{F}_g(R)$.

- $\varphi_n(\mathbb{F}_\theta) = \inf_{\hat{f}} \mathcal{R}_n^{(p)}[\hat{f}, \mathbb{F}_\theta]$ ($\mathbb{L}_p$–minimax risk)
  - $\mathcal{R}_n^{(p)}\left[\hat{f}, \mathbb{F}_\theta\right] := \sup_{f\in\mathbb{F}_\theta} \mathcal{R}_n^{(p)}[\hat{f}, f]$
  - Infimum is taken over all possible estimators.
- $(\theta, R) \in \Theta \times (1, \infty)$–nuisance parameter.
- $\psi_n(\mathbb{F}_\theta) = \sup_{f\in\mathbb{F}_\theta} \underbrace{\left\| \inf_{\vec{h}\in\mathbb{H}} \left\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n(\cdot, \vec{h}) \right\} \right\|_p}_{\text{main term in the r.h.s. of the oracle inequality}} + C_p n^{-1}.$
- To study the asymptotics $\psi_n(\boldsymbol{\Sigma}_\theta)\varphi_n^{-1}(\boldsymbol{\Sigma}_\theta)$, $\mathbf{n} \to \infty$

Since the construction of the estimator $\hat{f}_{\vec{h}(\cdot)}$ is independent of (θ, R) and p we can analyze its adaptivity under an arbitrary $\mathbb{L}_p$-loss over an arbitrary scale of functional classes $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$. But $\sup_{f\in\mathbb{F}_\theta} \left\| \inf_{\vec{h}\in\mathbb{H}} \left\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n(\cdot, \vec{h}) \right\} \right\|_p$ is not easy to analyze!

Adaptive minimax approach.

Let $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$ be a collection of subsets of $\mathbb{F}_{\boldsymbol{g}}(\boldsymbol{R})$.

- $\varphi_n(\mathbb{F}_\theta) = \inf_{\hat{f}} \mathcal{R}_n^{(p)}[\hat{f}, \mathbb{F}_\theta]$ ($\mathbb{L}_p$–minimax risk)
  - $\mathcal{R}_n^{(p)}[\hat{f}, \mathbb{F}_\theta] := \sup_{f\in\mathbb{F}_\theta} \mathcal{R}_n^{(p)}[\hat{f}, f]$
  - Infimum is taken over all possible estimators.
- $(\theta, R) \in \Theta \times (1, \infty)$–nuisance parameter.
- $\psi_n(\mathbb{F}_\theta) = \sup_{f\in\mathbb{F}_\theta} \underbrace{\Big\| \inf_{\vec{h}\in\mathbb{H}} \Big\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n(\cdot, \vec{h}) \Big\} \Big\|_p}_{\text{main term in the r.h.s. of the oracle inequality}} + C_p n^{-1}.$
- To study the asymptotics $\psi_n(\boldsymbol{\Sigma}_\theta)\varphi_n^{-1}(\boldsymbol{\Sigma}_\theta)$, $\mathbf{n} \to \infty$

Since the construction of the estimator $\widehat{f}_{\vec{\mathrm{h}}(\cdot)}$ is independent of (θ, R) and p we can analyze its adaptivity under an arbitrary $\mathbb{L}_p$-loss over an arbitrary scale of functional classes $\{\mathbb{F}_\theta\}_{\theta\in\Theta}$. But $\sup_{f\in\mathbb{F}_\theta} \Big\| \inf_{\vec{h}\in\mathbb{H}} \Big\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n(\cdot, \vec{h}) \Big\} \Big\|_p$ is not easy to analyze!

Objective.

- To bound from above

$$\sup_{f\in\mathbb{F}} \Big\| \inf_{\vec{h}\in\mathbb{H}} \Big\{ \mathcal{B}_{\vec{h}}(\cdot, f) + U_n\big(\cdot, \vec{h}\big) \Big\} \Big\|_p$$

$\forall \mathbb{F} \subset \mathbb{F}_{g,\mathbf{u}}(R, D) \cap \mathbb{B}_{\mathbf{q},d}(D)$, $\mathbf{q} \geq p$, $\mathbf{u} \geq \mathbf{q}$, $D > 0$.

- $\mathbb{F}_{g,\mathbf{u}}(R, D) := \Big\{ f \in \mathbb{F}_g(R) : \ \mathfrak{p} \in \mathbb{B}^{(\infty)}_{\mathbf{u},d}(D) \Big\}$;
- $\mathfrak{p} = (1-\alpha) f + \alpha [f \star g]$;
- $\mathbb{B}^{(\infty)}_{\mathbf{u},d}(D)$ denotes the ball in the weak-type $\mathbb{L}_{\mathbf{u},\infty}\big(\mathbb{R}^d\big)$:

$$\mathbb{B}^{(\infty)}_{\mathbf{u},d}(D) = \big\{ \lambda : \mathbb{R}^d \to \mathbb{R} : \|\lambda\|_{\mathbf{u},\infty} < D \big\};$$

$$\|\lambda\|_{\mathbf{u},\infty} = \inf \Big\{ C : \ \nu_d\big(x : |T(x)| > \mathfrak{z} \big) \leq C^{\mathbf{u}} \mathfrak{z}^{-\mathbf{u}}, \ \forall \mathfrak{z} > 0 \Big\}.$$

Important quantities.

- Quantities related to "approximation error" $\mathcal{B}_{\vec{h}}(\cdot, f)$:

Assumption 3.

$\mathcal{K}: \mathbb{R} \to \mathbb{R}$ is a compactly supported, bounded and $\int \mathcal{K} = 1$.

$$K(x) = \prod_{j=1}^{d} \mathcal{K}(x_j), \ \forall x \in \mathbb{R}^d$$

$$b_{\mathrm{v},f,j}(x) = \sup_{h \in \mathcal{H}:\, h \leq \mathrm{v}} \left| \int_{\mathbb{R}} \mathcal{K}(u) f\left(x + uh\mathbf{e}_j\right) \nu_1(\mathrm{d}u) - f(x) \right|;$$

- $(\mathbf{e}_1, \ldots, \mathbf{e}_d)$ denotes the canonical basis of $\mathbb{R}^d$.
- $\mathcal{H}$ is the diadic grid in $(0, \infty)$

$$\boxed{\mathrm{B}_{j,s,\mathbb{F}}(\mathrm{v}) = \sup_{f \in \mathbb{F}} \left\| b_{\mathrm{v},f,j} \right\|_s, \ \ s \in [1, \infty].}$$

Important quantities.

- Quantities related to "upper function" $U_n(\cdot, \vec{h})$:

$$F_n(\vec{h}) = \frac{\sqrt{\ln n + \sum_{j=1}^{d} |\ln h_j|}}{\sqrt{n} \prod_{j=1}^{d} h_j^{\frac{1}{2}} (h_j \wedge 1)^{\mu_j(\alpha)}};$$

$$G_n(\vec{h}) = \frac{\ln n + \sum_{j=1}^{d} |\ln h_j|}{n \prod_{j=1}^{d} h_j (h_j \wedge 1)^{\mu_j(\alpha)}}.$$

  - $\vec{\mu}(\alpha) = \vec{\mu}$ if $\alpha = 1$ and $\vec{\mu}(\alpha) = 0$ if $\alpha \neq 1$.

- Sets of bandwidths: for any $v, z > 0$

$$\mathfrak{H}(v) = \{\vec{h} \in \mathcal{H}^d : G_n(\vec{h}) \leq av\};$$

$$\mathfrak{H}(v, z) = \{\vec{h} \in \mathfrak{H}(v) : F_n(\vec{h}) \leq avz^{-1/2}\}.$$

  - $a > 0$ is explicitly known constant.

Important quantities.

- "Mixed" quantities: for any $\vec{s} \in [1,\infty)^d$, $u \geq 1$, $v > 0$

$$\Lambda_{\vec{s}}(v,\mathbb{F}) = \inf_{\vec{h}\in\mathfrak{H}(v)} \Big[\sum_{j=1}^{d} v^{-s_j} \big[\mathrm{B}_{j,s_j,\mathbb{F}}(h_j)\big]^{s_j} + v^{-2} F_n^2(\vec{h}) \Big];$$

$$\Lambda_{\vec{s}}(v,\mathbb{F},\mathrm{u}) = \inf_{z\geq 2} \inf_{\vec{h}\in\mathfrak{H}(v,z)} \Big[\sum_{j=1}^{d} v^{-s_j} \big[\mathrm{B}_{j,s_j,\mathbb{F}}(h_j)\big]^{s_j} + z^{-\mathrm{u}} \Big];$$

Define finally for any $0 < \underline{v} < \overline{v} \leq \infty$

$$\mathcal{I}_{\mathbb{F},\vec{s}}(\underline{v},\overline{v},u) = \int_{\underline{v}}^{\overline{v}} v^{p-1} \big[\Lambda_{\vec{s}}(v,\mathbb{F},\mathrm{u}) \wedge \Lambda_{\vec{s}}(v,\mathbb{F})\big] \mathrm{d}v$$

- "Tail" quantity: $\ell_{p,d}(v) = v^{p-1}(1+|\ln(v)|)^{d-1}$, $v > 0$.

Abstract Maximal Theorem

- Maximal risk of the p.s.r run over $\mathbb{H} = \mathcal{H}^d$

$$R_n(\mathbb{F}) = \sup_{f \in \mathbb{F}} \mathcal{R}_n^{(p)}\big[\widehat{f}_{\vec{\mathbf{h}}(\cdot)}; f\big]$$

Theorem 2. Let Assumptions 1, 2 and 3 be fulfilled. Then

For any $n \geq 3, p \in [1, \infty), R > 0, D > 0$, $\mathbf{q} \geq 1, \mathbf{u} \geq \mathbf{q}$, $0 < \underline{v} \leq \overline{v} \leq \infty$, $\vec{s} \in (1, \infty)^d$, $\vec{q} \in [p, \infty)^d$ and any $\mathbb{F} \subset \mathbb{B}_{\mathbf{q},d}(D) \cap \mathbb{F}_{g,\mathbf{u}}(R, D)$

$$R_n(\mathbb{F}) \leq A\big[\ell_{p,d}(\underline{v}) + \mathcal{I}_{\mathbb{F},\vec{s}}(\underline{v}, \overline{v}, u) + \overline{v}^p \Lambda_{\vec{q}}(\overline{v}, \mathbb{F}, \mathbf{u})\big]^{\frac{1}{p}} + Cn^{-1}.$$

If additionally $q > p$ then

$$R_n(\mathbb{F}) \leq A\big[\ell_{p,d}(\underline{v}) + \mathcal{I}_{\mathbb{F},\vec{s}}(\underline{v}, \overline{v}, u) + \overline{v}^{p-q}\big]^{\frac{1}{p}} + Cn^{-1}.$$

At last if $q = \infty$ then

$$R_n(\mathbb{F}) \leq A\big[\ell_{p,d}(\underline{v}) + \mathcal{I}_{\mathbb{F},\vec{s}}(\underline{v}, \overline{v}, u) + \Lambda_{\vec{s}}(\overline{v}, \mathbb{F}, \mathbf{u})\big]^{\frac{1}{p}} + Cn^{-1}.$$

- C depends only on $g, \mathcal{K}, p, d$.
- A depends only on $g, R, D, \mathcal{K}, p, d, u, q, \vec{s}, \vec{q}$.

Adaptive estimation over the scale of anisotropic Nikol'skii classes

I^0. Case $\alpha = 0$ (classical density model)

$$\mathbb{F}_{\theta} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \ \theta = (\vec{\beta}, \vec{L}, \vec{r}).$$

- Full characterization of minimax rates. We discovered **7**! different regimes of the asymptotics of minimax risk, including inconsistency zone.
- We proved that our pointwise selection rule leads to optimally-adaptive (in some regimes) or nearly optimally-adaptive estimator (up to logarithmic factor).

II^0. Case $\alpha = 1$ (density deconvolution)

$$\mathbb{F}_{\theta} = \mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}), \ \theta = (\vec{\beta}, \vec{L}, \vec{r}).$$

- Under additional assumption $\|g\|_\infty < \infty$ we obtained full characterization of minimax rates (**5** regimes, including inconsistency zone). Also we prove that our estimator is optimally or nearly-optimally adaptive.

Adaptive estimation over the scale of anisotropic Nikol'skii classes

III^{0}. Case $\alpha \in [0, 1]$ (convolution structure density model)

$$\mathbb{F}_{\theta} = \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big) \cap \mathbb{B}_{\infty,d}(D), \;\; \theta = \big(\vec{\beta}, \vec{L}, \vec{r}, D\big).$$

- We obtained full characterization of minimax rates (**4** regimes). Also we prove that our estimator is optimally or nearly-optimally adaptive.

Adaptive estimation. Unbounded case, $\alpha = 1$.

- Collection of classes: $\mathbf{u} = \infty \Leftrightarrow \|g \star f\|_\infty \le D$.

$$\mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}, R, D\big) := \mathbb{N}_{\vec{r},d}\big(\vec{\beta}, \vec{L}\big) \cap \mathbb{F}_{g,\infty}(R, D).$$

- Nuisance parameter: $\big(\vec{\beta}, \vec{r}, \vec{L}, R, D\big)$
- Important quantities: for any $s \in [1, \infty]$ and $\alpha \in [0, 1]$

$\tau(s) = 1 - 1/\omega(0) + 1/(\beta(0)s)$ $\kappa_\alpha(s) = \omega(\alpha)\big(2 + 1/\beta(\alpha)\big) - s$

- $\frac{1}{\beta(\alpha)} := \sum_{j=1}^{d} \frac{2\mu_j(\alpha)+1}{\beta_j}, \quad \frac{1}{\omega(\alpha)} := \sum_{j=1}^{d} \frac{2\mu_j(\alpha)+1}{\beta_j r_j}$.
- $\vec{\mu}(\alpha) = \vec{\mu}$, $\alpha = 1$, $\vec{\mu}(\alpha) = (0, \dots, 0)$, $\alpha \in [0, 1)$.
- $p^* = p \vee \max_{j=1,\dots,d} r_j$.

Adaptive estimation. Unbounded case, $\alpha = 1$.

- Important quantities: for any $s \in [1, \infty]$ and $\alpha \in [0, 1]$

$$\tau(s) = 1 - 1/\omega(0) + 1/(\beta(0)s) \qquad \kappa_\alpha(s) = \omega(\alpha)\big(2 + 1/\beta(\alpha)\big) - s$$

  - $\frac{1}{\beta(\alpha)} := \sum_{j=1}^{d} \frac{2\mu_j(\alpha)+1}{\beta_j}, \quad \frac{1}{\omega(\alpha)} := \sum_{j=1}^{d} \frac{2\mu_j(\alpha)+1}{\beta_j r_j}.$
  - $\vec{\mu}(\alpha) = \vec{\mu}$, $\alpha = 1$, $\vec{\mu}(\alpha) = (0, \dots, 0)$, $\alpha \in [0, 1)$.
  - $p^* = p \vee \max_{j=1,\dots,d} r_j$.

- Different regimes of the behavior of the minimax risk corresponds to the following relations

$$\kappa_\alpha(p) > p\omega(\alpha), \qquad 0 < \kappa_\alpha(p) \le p\omega(\alpha);$$

$$\kappa_\alpha(p) \le 0, \tau(p^*) > 0, \qquad \kappa_\alpha(p) \le 0, \tau(p^*) \le 0 \; p^* > p$$

■-tail zone, ■-dense zone, ■-sparse zone 1, ■-sparse zone 2.

■-inconsistency zone: $\kappa_\alpha(p) \le 0, \tau(p^*) \le 0 \; p^* = p$.

Adaptive upper bound, $\alpha = 1$.

- If $\kappa_\alpha(p) \le 0$ (the sparse zone)

$$\sup_{n\ge 1} \frac{\psi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},R,D)\big)}{\varphi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},R,D)\big)} \le C < \infty.$$

We assert that the proposed estimator is optimally adaptive on the whole sparse zone.

- If $0 < \kappa_\alpha(p) \le p\omega(\alpha)$ (the dense zone)

$$\frac{\psi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},R,D)\big)}{\varphi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},R,D)\big)} \le C(\ln n)^{\rho(\alpha)}, \quad \forall n \ge 1$$

- If $\kappa_\alpha(p) > p\omega(\alpha)$ (the tail zone)

$$\frac{\psi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},R,D)\big)}{\varphi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta},\vec{L},R,D)\big)} \le C(\ln n)^{\rho(\alpha)+\frac{d-1}{p}}, \quad \forall n \ge 1$$

We assert that the proposed estimator is nearly-optimally adaptive on the dense and tail zones.

Lower bound in unbounded case. $\|g\|_\infty < \infty$.

For any $(\vec{\beta}, \vec{r}, \vec{L}, R, D)$

$$\boxed{\varphi_n\big(\mathbb{N}_{\vec{r},d}(\vec{\beta}, \vec{L}, R, M)\big) \gtrsim \delta_n^{\rho(\alpha)}}$$

■-tail zone, ■-dense zone, ■-sparse zone 1, ■-sparse zone 2.

$$\rho(\alpha) = \begin{cases} \frac{1-1/p}{1-1/\omega(\alpha)+1/\beta(\alpha)}, & \kappa_\alpha(p) > p\omega(\alpha); \\ \frac{\beta(\alpha)}{2\beta(\alpha)+1}, & 0 < \kappa_\alpha(p) \le p\omega(\alpha); \\ \frac{\tau(p)}{(2+1/\beta(\alpha))\tau(\infty)+1/[\omega(\alpha)\beta(0)]}, & \kappa_\alpha(p) \le 0, \quad \tau(p^*) > 0; \\ \frac{\omega(\alpha)(1-p^*/p)}{\kappa_\alpha(p^*)}, & \kappa_\alpha(p) \le 0, \; \tau(p^*) \le 0. \end{cases}$$

$$\delta_n = \begin{cases} n^{-1}, & \kappa_\alpha(p) > 0; \\ n^{-1}\ln(n), & \kappa_\alpha(p) \le 0. \end{cases}$$

Anisotropic Nikolskii classes

Let $(e_1,\ldots,e_d)$ denote the canonical basis of $\mathbb{R}^d$. For function $g:\mathbb{R}^d\to\mathbb{R}^1$ and real number $u\in\mathbb{R}$ define *the first order difference operator with step size u in direction of the variable x_j*

$$\Delta_{u,j}g(x)=g(x+ue_j)-g(x),\quad j=1,\ldots,d.$$

The **k**-th order difference operator is defined as

$$\Delta^k_{u,j}g(x)=\Delta_{u,j}\Delta^{k-1}_{u,j}g(x)=\sum_{l=1}^k(-1)^{l+k}\binom{k}{l}\Delta_{ul,j}g(x).$$

Definition

For given numbers $\vec{r}=(r_1,\ldots,r_d)\in[1,\infty]^d$, $\vec{\beta}=(\beta_1,\ldots,\beta_d)\in(0,\infty)^d$ and $\vec{L}=(L_1,\ldots,L_d)\in(0,\infty)^d$ we say that $g:\mathbb{R}^d\to\mathbb{R}^1$ belongs to $\mathbb{N}_{\vec{r},d}\big(\vec{\beta},\vec{L}\big)$ if

- $\|g\|_{r_j}\leq L_j,\quad \forall j=\overline{1,d};$
- $\forall j=\overline{1,d}\ \ \exists k_j>\beta_j$ such that

$$\Big\|\Delta^{k_j}_{u,j}g\Big\|_{r_j}\leq L_j|u|^{\beta_j},\quad \forall u\in\mathbb{R}^d,\quad \forall j=\overline{1,d}.$$