# Uniform bounds for positive random functionals with application to density estimation

Oleg Lepski

**Laboratoire d'Analyse, Topologie et Probabilités**
**Université de Provence**

Congrès SMAI 2011, May 23 - May 27, Guidel

# Outline

1. Upper functions. General case
2. Probabilistic study of statistical objects
   - Upper functions. Special cases
   - Comparison with asymptotical results
3. Sup-norm oracle inequality in density estimation
   - Selection from the family of kernel estimators
   - Adaptation over anisotropic Hölder classes

# Part I

## Upper functions. General case

Let $(\Omega, \mathfrak{A}, \mathrm{P})$ be a probability space, $\mathfrak{S}$ be a linear space and $\Theta$ be a given set.

Let $\chi_n : \Theta \times \Omega \to \mathfrak{S}, n \in \mathbb{N}^*$, be a given sequence of $\mathfrak{A}$-measurable maps and $\mathbb{P}_f^{(n)}$ be the corresponding sequence of probability laws, parameterized by $f \in \mathbb{F}$.

Let $\Psi : \mathfrak{S} \to \mathbb{R}_+$ be a given *sub-additive* functional.

**Goal: find non-random positive function on $\Theta$ which would be uniform upper bound for $\Psi(\chi_{n,\theta})$ in the sense**

$$\mathbb{P}_f^{(n)}\Big\{\sup_{\theta\in\Theta}\big[\Psi\big(\chi_{n,\theta}\big) - \mathcal{U}_n(y,\theta)\big] \geq 0\Big\} \leq \mathcal{P}_n(y,f);$$

$$\mathbb{E}_f^{(n)}\Big(\sup_{\theta\in\Theta}\big[\Psi\big(\chi_{n,\theta}\big) - U_n(y,\theta)\big]\Big)_+^q \leq \mathcal{E}_n(y,f,q).$$

The quantities $\mathcal{P}_n(y,f)$ and $\mathcal{E}_n(y,f,q)$ should possess several properties discussed later. In particular for any **fixed** $y > y_q$
$\mathcal{E}_n(y,f,q) \to 0, \quad n \to \infty$ **uniformly** w.r.t. $f$.

# Statistical models. Adaptive estimation.

**Regression model**

$$\mathbf{Y_i = f(z_i) + \xi_i, \quad i = \overline{1, n}}$$

$\xi_i$, $i \in \mathbb{N}^*$ are i.i.d.: $\mathbb{E}\xi_1 = 0$, or $\mathbf{med}(\xi_1) = 0$;

The design points $z_i \in \mathbb{R}^d$, $i = \overline{1, n}$ are supposed to be either fixed real vectors or i.i.d. random vectors.

We observe $X^{(n)} = \{(Y_1, z_1), \dots, (Y_n, z_n)\}$.

**Density model**

$$X^{(n)} \quad \sim \quad p_n(x) = \prod_{i=1}^{n} f(x_i), \quad x = (x_1, \dots, x_d) \in \mathbb{R}^d$$

$X^{(n)} = (X_1, \dots, X_n)$, $n \in \mathbb{N}^*$, where $X_i \in \mathbb{R}^d$, $i \in \mathbb{N}^*$, are i.i.d. random vectors having the density $f$.

**Goal:** **to estimate the function $f$ from the observation $X^{(n)}$.**

# Statistical models. Adaptive estimation.

**Regression model**

$$Y_i = f(z_i) + \xi_i, \quad i = \overline{1, n}$$

$\xi_i$, $i \in \mathbb{N}^*$ are i.i.d.: $\mathbb{E}\xi_1 = 0$, or $\mathrm{med}(\xi_1) = 0$;
The design points $z_i \in \mathbb{R}^d$, $i = \overline{1, n}$ are supposed to be either fixed real vectors or i.i.d. random vectors.
We observe $X^{(n)} = \{(Y_1, z_1), \ldots, (Y_n, z_n)\}$.

**Density model**

$$X^{(n)} \quad \sim \quad p_n(x) = \prod_{i=1}^{n} f(x_i), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d$$

$X^{(n)} = (X_1, \ldots, X_n), n \in \mathbb{N}^*$, where $X_i \in \mathbb{R}^d$, $i \in \mathbb{N}^*$, are i.i.d. random vectors having the density $f$.

Goal: **to estimate the function $f$ from the observation $X^{(n)}$.**

Let $(\Omega, \mathfrak{A}, \mathrm{P})$ be a probability space, $\mathfrak{S}$ be a linear space and $\Theta$ be a given set.

Let $\chi_n : \Theta \times \Omega \to \mathfrak{S}, n \in \mathbb{N}^*$, be a given sequence of $\mathfrak{A}$-measurable maps and $\mathbb{P}_f^{(n)}$ be the corresponding sequence of probability laws.

Let $\Psi : \mathfrak{S} \to \mathbb{R}_+$ be a given *sub-additive* functional.

**Goal: find non-random positive function on $\theta$ which would be uniform upper bound for $\Psi(\chi_{n,\theta})$ in the sense**

$$\mathbb{P}_f^{(n)}\Big\{ \sup_{\theta\in\Theta} \big[\Psi\big(\chi_{n,\theta}\big) - \mathcal{U}_n(y,\theta)\big] \geq 0\Big\} \leq \mathcal{P}_n(y,f);$$

$$\mathbb{E}_f^{(n)}\Big( \sup_{\theta\in\Theta} \big[\Psi\big(\chi_{n,\theta}\big) - \mathrm{U}_n(y,\theta)\big]\Big)_+^q \leq \mathcal{E}_n(y,f,q).$$

To realize this program we impose the Bernstein-type assumption on the tail probability for $\Psi\big(\chi_{n,\theta}\big)$ for any given $\theta \in \Theta$ and $\Psi\big(\chi_{n,\theta_1} - \chi_{n,\theta_2}\big),\ \theta_1, \theta_2 \in \Theta$.

Furthermore $\mathbb{P} = \mathbb{P}_f^{(n)}$, $\mathbb{E} = \mathbb{E}_f^{(n)}$ and $\chi_{\theta} = \chi_{n,\theta}$.

**Assumption (1)**

1. *There exist* $\mathbf{A}, \mathbf{B} : \mathbf{\Theta} \to \mathbb{R}_+$ *such that* $\forall \theta \in \mathbf{\Theta}$ *and* $\forall \mathbf{z} > \mathbf{0}$

$$\mathbb{P}\left\{\mathbf{\Psi}(\chi_{\theta}) \geq \mathbf{z}\right\} \leq \mathbf{G}\left\{\frac{\mathbf{z}^2}{\mathbf{A}^2(\theta) + \mathbf{B}(\theta)\mathbf{z}}\right\}$$

2. *There exist* $\mathbf{a}, \mathbf{b} : \mathbf{\Theta} \times \mathbf{\Theta} \to \mathbb{R}_+$ *s.t.* $\forall \theta_1, \theta_2 \in \mathbf{\Theta}$, $\forall \mathbf{z} > \mathbf{0}$

$$\mathbb{P}\left\{\mathbf{\Psi}(\chi_{\theta_1} - \chi_{\theta_2}) \geq \mathbf{z}\right\} \leq \mathbf{G}\left\{\frac{\mathbf{z}^2}{\mathbf{a}^2(\theta_1, \theta_2) + \mathbf{b}(\theta_1, \theta_2)\mathbf{z}}\right\}$$

$\mathbf{G}(\mathbf{x}) = \mathbf{c}\exp\{-\mathbf{x}\}$, $\mathbf{c} > 0$.

Furthermore $\mathbb{P} = \mathbb{P}_f^{(n)}$, $\mathbb{E} = \mathbb{E}_f^{(n)}$ and $\chi_{\theta} = \chi_{n,\theta}$.

**Assumption (1)**

1. *There exist* $\mathbf{A}, \mathbf{B} : \boldsymbol{\Theta} \to \mathbb{R}_+$ *such that* $\forall \theta \in \boldsymbol{\Theta}$ *and* $\forall \mathbf{z} > \mathbf{0}$
$$\mathbb{P}\left\{\boldsymbol{\Psi}(\chi_\theta) \geq \mathbf{z}\right\} \leq \mathbf{G}\left\{\frac{\mathbf{z}^2}{\mathbf{A}^2(\theta) + \mathbf{B}(\theta)\mathbf{z}}\right\}$$
2. *There exist* $\mathbf{a}, \mathbf{b} : \boldsymbol{\Theta} \times \boldsymbol{\Theta} \to \mathbb{R}_+$ *s.t.* $\forall \theta_1, \theta_2 \in \boldsymbol{\Theta}$, $\forall \mathbf{z} > \mathbf{0}$
$$\mathbb{P}\left\{\boldsymbol{\Psi}(\chi_{\theta_1} - \chi_{\theta_2}) \geq \mathbf{z}\right\} \leq \mathbf{G}\left\{\frac{\mathbf{z}^2}{\mathbf{a}^2(\theta_1, \theta_2) + \mathbf{b}(\theta_1, \theta_2)\mathbf{z}}\right\}$$

$\mathbf{G}(\mathbf{x}) = c\,\mathbf{exp}\left\{-\mathbf{x}\right\}$, $c > 0$.

# General case

**Assumptions. Bound for a given trajectory.** $G(x) = c\exp\{-x\}$, $c > 0$.

## Assumption (1)

1. *There exist* $\mathbf{A}, \mathbf{B} : \Theta \to \mathbb{R}_+$ *such that* $\forall \theta \in \Theta$ *and* $\forall z > 0$

$$\mathbb{P}\{\Psi(\chi_\theta) \geq z\} \leq G\left\{\frac{z^2}{A^2(\theta) + B(\theta)z}\right\}$$

2. *There exist* $\mathbf{a}, \mathbf{b} : \Theta \times \Theta \to \mathbb{R}_+$ *s.t.* $\forall \theta_1, \theta_2 \in \Theta$, $\forall z > 0$

$$\mathbb{P}\{\Psi(\chi_{\theta_1} - \chi_{\theta_2}) \geq z\} \leq G\left\{\frac{z^2}{a^2(\theta_1, \theta_2) + b(\theta_1, \theta_2)z}\right\}$$

## Assumption (2)

1. *The mappings* $\mathbf{a}$ *and* $\mathbf{b}$ *are semi-metrics on* $\Theta$ *and* $\chi_\bullet$ *is stochastically continuous in the topology generated by* $\mathbf{a} \vee \mathbf{b}$.
2. $\Theta$ *is totally bounded with respect to the semi-metric* $\mathbf{a} \vee \mathbf{b}$ *and* $\overline{A}_\theta := \sup_{\theta\in\Theta} A(\theta) < \infty$, $\overline{B}_\theta := \sup_{\theta\in\Theta} B(\theta) < \infty$.

Let $\mathbf{X}$ be $\mathcal{X}$-valued random vector defined on $(\Omega, \mathfrak{A}, \mathrm{P})$ and let $\mathbf{X_i}, \mathbf{i} = \overline{\mathbf{1}, \mathbf{n}}$ be independent copies of $\mathbf{X}$.
Let $\mathbf{W}$ be a given set of functions $\mathbf{w} : \boldsymbol{\mathcal{X}} \to \mathbb{R}$.

**Example: $\Psi(\chi_\theta) = \left|\mathfrak{D}_w(\cdot)\right|, \quad \theta = w, \quad \Theta = W$**

$$\mathfrak{D}_w = \sum_{i=1}^{n} \left[w(X_i) - \mathbb{E}w(X)\right].$$

If $\mathbf{W}$ is a subset of the set of bounded functions then **Assumption 1** follows from Bernstein inequality. Here

$$A(w) = \sqrt{\mathbb{E}\left[w(X)\right]^2}, \qquad B(w) = \sup_{x \in \mathcal{X}} \left|w(x)\right|;$$

$$a\left(w_1, w_2\right) = A\left(w_1 - w_2\right), \qquad b\left(w_1, w_2\right) = B\left(w_1 - w_2\right).$$

We remark that **Assumption 2 (1)** is also fulfilled.

**Example: $\Psi(\chi_\theta) = \left|\chi_\theta\right|$ , $\chi_\theta$ is zero mean gaussian function**

The **Assumptions 1 and 2 (1)** are obviously fulfilled with
$\mathbf{B} = \mathbf{b} \equiv 0$

$$\mathbf{A}(\theta) = \sqrt{\mathbb{E}\big(\chi_\theta\big)^2},\ \mathbf{a}(\theta_1, \theta_2) = \sqrt{\mathbb{E}\big(\chi_{\theta_1} - \chi_{\theta_2}\big)^2}$$

# General case: bounds under Assumptions 1 and 2.

**Assumption (1)**

1. *There exist* $\mathbf{A}, \mathbf{B} : \mathbf{\Theta} \to \mathbb{R}_+$ *such that* $\forall \theta \in \mathbf{\Theta}$ *and* $\forall \mathbf{z} > \mathbf{0}$
$$\mathbb{P}\left\{\Psi(\chi_\theta) \geq \mathbf{z}\right\} \leq \mathbf{G}\left\{\frac{\mathbf{z}^2}{\mathbf{A}^2(\theta) + \mathbf{B}(\theta)\mathbf{z}}\right\}$$
2. *There exist* $\mathbf{a}, \mathbf{b} : \mathbf{\Theta} \times \mathbf{\Theta} \to \mathbb{R}_+$ *s.t.* $\forall \theta_1, \theta_2 \in \mathbf{\Theta}$, $\forall \mathbf{z} > \mathbf{0}$
$$\mathbb{P}\left\{\Psi(\chi_{\theta_1} - \chi_{\theta_2}) \geq \mathbf{z}\right\} \leq \mathbf{G}\left\{\frac{\mathbf{z}^2}{\mathbf{a}^2(\theta_1, \theta_2) + \mathbf{b}(\theta_1, \theta_2)\mathbf{z}}\right\}$$

**Assumption (2)**

1. *The mappings* $\mathbf{a}$ *and* $\mathbf{b}$ *are semi-metrics on* $\mathbf{\Theta}$ *and* $\chi_\bullet$ *is stochastically continuous in the topology generated by* $\mathbf{a} \vee \mathbf{b}$.
2. $\mathbf{\Theta}$ *is totally bounded with respect to the semi-metric* $\mathbf{a} \vee \mathbf{b}$ *and* $\overline{\mathbf{A}}_\theta := \sup_{\theta \in \Theta} \mathbf{A}(\theta) < \infty$, $\overline{\mathbf{B}}_\theta := \sup_{\theta \in \Theta} \mathbf{B}(\theta) < \infty$.

# General case: bounds under Assumptions 1 and 2.

The most important elements of our construction are:

$$\Theta_A(t) = \Big\{\theta \in \Theta : \ A(\theta) \le t\Big\}, \quad t > 0;$$

$$\Theta_B(t) = \Big\{\theta \in \Theta : \ B(\theta) \le t\Big\}, \quad t > 0;$$

For any $x > 0$, any $\widetilde{\Theta} \subseteq \Theta$ and any $s \in \mathbb{S}$

$$e_s^{(a)}\big(x, \widetilde{\Theta}\big) = \sup_{\delta>0} \delta^{-2} \mathfrak{E}_{\widetilde{\Theta},\, \mathrm{a}}\big(x(48\delta)^{-1}s(\delta)\big)$$

$$e_s^{(b)}\big(x, \widetilde{\Theta}\big) = \sup_{\delta>0} \delta^{-1} \mathfrak{E}_{\widetilde{\Theta},\, \mathrm{b}}\big(x(48\delta)^{-1}s(\delta)\big)$$

- $\mathfrak{E}_{\widetilde{\Theta},\, \mathrm{d}}(\nu),\ \nu > 0,$ - entropy of $\widetilde{\Theta}$ measured in semi-metric $\mathbf{d}$;
- $\mathbb{S} = \Big\{s : \mathbb{R} \to \mathbb{R}_+ \setminus \{0\} : \ \sum_{k=0}^{\infty} s\big(2^{k/2}\big) \le 1\Big\}.$

Introduced quantities

$$\Theta_A(t) = \Big\{\theta \in \Theta : \; A(\theta) \le t\Big\}, \quad t > 0;$$

$$\Theta_B(t) = \Big\{\theta \in \Theta : \; B(\theta) \le t\Big\}, \quad t > 0;$$

$$e_s^{(a)}\big(x, \widetilde{\Theta}\big) = \sup_{\delta>0} \delta^{-2} \mathfrak{E}_{\widetilde{\Theta},\, a}\big(x(48\delta)^{-1}s(\delta)\big)$$

$$e_s^{(b)}\big(x, \widetilde{\Theta}\big) = \sup_{\delta>0} \delta^{-1} \mathfrak{E}_{\widetilde{\Theta},\, b}\big(x(48\delta)^{-1}s(\delta)\big)$$

allow us to define for any $u, v \ge 1$ and any $\vec{s} = (s_1, s_2)$

$$\widetilde{\mathcal{E}}_{\vec{s}}(u,v) = e_{s_1}^{(a)}\Big(\underline{A}u, \Theta_A\big(\underline{A}u\big)\Big) + e_{s_2}^{(b)}\Big(\underline{B}v, \Theta_B\big(\underline{B}v\big)\Big)$$

$$\underline{A} = \inf_{\theta\in\Theta} A(\theta) > 0; \quad \underline{B} = \inf_{\theta\in\Theta} B(\theta) > 0.$$

Example: $s_1 = s_2 = (6/\pi^2)\big(1 + [\ln x]^2\big)^{-1}, \; x \ge 0$

# General case: bounds under Assumptions 1 and 2.

Introduced quantities

$$\Theta_A(t) = \left\{\theta \in \Theta : \ A(\theta) \le t\right\}, \quad t > 0;$$

$$\Theta_B(t) = \left\{\theta \in \Theta : \ B(\theta) \le t\right\}, \quad t > 0;$$

$$e_s^{(a)}\left(x, \widetilde{\Theta}\right) = \sup_{\delta>0} \delta^{-2} \mathfrak{E}_{\widetilde{\Theta},\, a}\left(x(48\delta)^{-1} s(\delta)\right)$$

$$e_s^{(b)}\left(x, \widetilde{\Theta}\right) = \sup_{\delta>0} \delta^{-1} \mathfrak{E}_{\widetilde{\Theta},\, b}\left(x(48\delta)^{-1} s(\delta)\right)$$

allow us to define for any $u, v \ge 1$ and any $\vec{s} = (s_1, s_2)$

$$\widetilde{\mathcal{E}}_{\vec{s}}(u, v) = e_{s_1}^{(a)}\left(\underline{A}u, \Theta_A\left(\underline{A}u\right)\right) + e_{s_2}^{(b)}\left(\underline{B}v, \Theta_B\left(\underline{B}v\right)\right)$$

$$\underline{A} = \inf_{\theta\in\Theta} A(\theta) > 0; \quad \underline{B} = \inf_{\theta\in\Theta} B(\theta) > 0.$$

**Example**: $s_1 = s_2 = (6/\pi^2)\left(1 + [\ln x]^2\right)^{-1}, \ x \ge 0$

# General case: bounds under Assumptions 1 and 2.

Denote $\ell(\mathbf{u}) = \ln\{1 + \ln(\mathbf{u})\} + 2\ln\{1 + \ln\{1 + \ln(\mathbf{u})\}\}$ and set for any $\theta \in \Theta$ and $\varepsilon > 0, \mathbf{q} \geq 0$

"Probability payment"

$$\mathsf{P}_\varepsilon(\theta) = 2\left[1 + \varepsilon^{-1}\right]^2 \mathcal{E}\big(\mathcal{A}_\varepsilon(\theta), \mathcal{B}_\varepsilon(\theta)\big) + \left[\ell\big(\mathcal{A}_\varepsilon(\theta)\big) + \ell\big(\mathcal{B}_\varepsilon(\theta)\big)\right]$$

$$\mathcal{A}_\varepsilon(\theta) = (1+\varepsilon)\left[\mathsf{A}(\theta)/\underline{\mathsf{A}}\right], \qquad \mathcal{B}_\varepsilon(\theta) = (1+\varepsilon)\left[\mathsf{B}(\theta)/\underline{\mathsf{B}}\right]$$

"Moment payment"

$$\mathsf{M}_{\varepsilon,\mathbf{q}}(\theta) = 2\left[1 + \varepsilon^{-1}\right]^2 \mathcal{E}\big(\mathcal{A}_\varepsilon(\theta), \mathcal{B}_\varepsilon(\theta)\big) + (\varepsilon + \mathbf{q})\ln\left[\mathcal{A}_\varepsilon(\theta)\mathcal{B}_\varepsilon(\theta)\right]$$

**Remark:** $\mathcal{E}$ is an arbitrary function satisfying $\widetilde{\mathcal{E}}_{\vec{\mathbf{s}}}(\cdot,\cdot) \leq \mathcal{E}(\cdot,\cdot)$.

**Remark:** $\varepsilon$ and $\vec{\mathbf{s}}$ are turning parameters.

# General case: bounds under Assumptions 1 and 2.

Denote $\ell(u) = \ln\{1+\ln(u)\} + 2\ln\{1+\ln\{1+\ln(u)\}\}$ and set for any $\theta \in \Theta$ and $\varepsilon > 0, q \geq 0$

**"Probability payment"**

$$P_\varepsilon(\theta) = 2\big[1+\varepsilon^{-1}\big]^2 \mathcal{E}\big(\mathcal{A}_\varepsilon(\theta), \mathcal{B}_\varepsilon(\theta)\big) + \big[\ell\big(\mathcal{A}_\varepsilon(\theta)\big) + \ell\big(\mathcal{B}_\varepsilon(\theta)\big)\big]$$

$$\mathcal{A}_\varepsilon(\theta) = (1+\varepsilon)\big[A(\theta)/\underline{A}\big], \qquad \mathcal{B}_\varepsilon(\theta) = (1+\varepsilon)\big[B(\theta)/\underline{B}\big]$$

**"Moment payment"**

$$M_{\varepsilon,q}(\theta) = 2\big[1+\varepsilon^{-1}\big]^2 \mathcal{E}\big(\mathcal{A}_\varepsilon(\theta), \mathcal{B}_\varepsilon(\theta)\big) + (\varepsilon + q)\ln\big[\mathcal{A}_\varepsilon(\theta)\mathcal{B}_\varepsilon(\theta)\big]$$

**Remark:** $\mathcal{E}$ is an arbitrary function satisfying $\widetilde{\mathcal{E}}_{\vec{s}}(\cdot,\cdot) \leq \mathcal{E}(\cdot,\cdot)$.

**Remark:** $\varepsilon$ and $\vec{s}$ are turning parameters.

## UPPER FUNCTIONS OF THE FIRST TYPE ($\underline{A}, \underline{B} > 0$)

1. "Probability upper function" $\rightsquigarrow$ $V^{(z,\varepsilon)}(\theta)$

$$(1+\varepsilon)^4 \left( A(\theta)\sqrt{P_\varepsilon(\theta) + (1+\varepsilon)^2 z} + B(\theta)\left[P_\varepsilon(\theta) + (1+\varepsilon)^2 z\right]\right)$$

2. "Moment's upper function" $\rightsquigarrow$ $U_q^{(z,\varepsilon)}(\theta)$

$$(1+\varepsilon)^4 \left( A(\theta)\sqrt{M_{\varepsilon,q}(\theta) + (1+\varepsilon)^2 z} + B(\theta)\left[M_{\varepsilon,q}(\theta) + (1+\varepsilon)^2 z\right]\right)$$

# General case: bounds under Assumptions 1 and 2.

$$V^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4 \left( A(\theta)\sqrt{P_\varepsilon(\theta) + (1+\varepsilon)^2 z} + B(\theta)\left[P_\varepsilon(\theta) + (1+\varepsilon)^2 z\right]\right)$$

$$U_q^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4 \left( A(\theta)\sqrt{M_{\varepsilon,q}(\theta) + (1+\varepsilon)^2 z} + B(\theta)\left[M_{\varepsilon,r}(\theta) + (1+\varepsilon)^2 z\right]\right)$$

**Proposition 1.** $\forall \vec{s} \in \mathbb{S} \times \mathbb{S}$, $\forall \varepsilon \in \left(0, \sqrt{2}-1\right]$, $\forall z \geq 1$

$$\mathrm{P}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - V^{(z,\varepsilon)}(\theta)\right] \geq 0\right\} \leq C_\varepsilon \exp\left\{-z\right\};$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - U_q^{(z,\varepsilon)}(\theta)\right]\right\}_+^q \leq C_{\varepsilon,q}\left[\underline{A} \vee \underline{B}\right]^q \exp\left\{-z\right\}.$$

$$C_\varepsilon = 2c\left[1 + \left[\ln\left\{1 + \ln\left(1+\varepsilon\right)\right\}\right]^{-2}\right]^2;$$

$$C_{\varepsilon,q} = c2^{(5q/2)+2}\Gamma(q+1)\,\varepsilon^{-q-4}.$$

**Assumption (1)**

1. *There exist* $\mathbf{A}, \mathbf{B} : \mathbf{\Theta} \to \mathbb{R}_+$ *such that* $\forall \theta \in \mathbf{\Theta}$ *and* $\forall \mathbf{z} > \mathbf{0}$

$$\mathbb{P}\left\{\mathbf{\Psi}(\chi_\theta) \geq \mathbf{z}\right\} \leq \mathrm{c}\,\mathbf{exp}\left\{-\frac{\mathbf{z}^2}{\mathbf{A}^2(\theta) + \mathbf{B}(\theta)\mathbf{z}}\right\}$$

It is equivalent to: $\forall \theta \in \mathbf{\Theta}, \forall \mathbf{z} \geq \mathbf{0}$ and $\forall \mathbf{q} \geq \mathbf{0}$

$$\mathbb{P}\left\{\mathbf{\Psi}(\chi_\theta) \geq \mathbf{A}(\theta)\sqrt{\mathbf{z}} + \mathbf{B}(\theta)\mathbf{z}\right\} \leq \mathrm{c}\,\mathbf{exp}\left\{-\mathbf{z}\right\},$$

$$\mathbb{E}\left\{\mathbf{\Psi}(\chi_\theta) - \left[\mathbf{A}(\theta)\sqrt{\mathbf{z}} + \mathbf{B}(\theta)\mathbf{z}\right]\right\}_+^{\mathbf{q}} \leq \mathbf{c_q}\left[\mathbf{A}(\theta) \vee \mathbf{B}(\theta)\right]^{\mathbf{q}} \mathbf{exp}\left\{-\mathbf{z}\right\}$$

$$\mathbf{c_q} = \mathrm{c}2^{\mathbf{q}}\mathbf{\Gamma}(\mathbf{q}+1).$$

Thus, the function $\mathbf{U}^{(\mathbf{z})}(\theta) := \mathbf{A}(\theta)\sqrt{\mathbf{z}} + \mathbf{B}(\theta)\mathbf{z}$ can be viewed as "pointwise upper function" for $\mathbf{\Psi}$, i.e. for fixed $\theta$.

# Bounds in general case: Payment for uniformity

$$\mathbb{P}\Big\{\Psi(\chi_\theta) \geq \mathsf{A}(\theta)\sqrt{z} + \mathsf{B}(\theta)z\Big\} \leq c\exp\{-z\},$$

$$\mathbb{E}\Big\{\Psi(\chi_\theta) - \Big[\mathsf{A}(\theta)\sqrt{z} + \mathsf{B}(\theta)z\Big]\Big\}_+^q \leq c_q\Big[\mathsf{A}(\theta) \vee \mathsf{B}(\theta)\Big]^q \exp\{-z\}$$

**Proposition 1.** $\forall \vec{s} \in \mathbb{S} \times \mathbb{S}$, $\forall \varepsilon \in \big(0, \sqrt{2} - 1\big]$, $\forall z \geq 1$

$$\mathrm{P}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{V}^{(z,\varepsilon)}(\theta)\right] \geq 0\right\} \leq C_\varepsilon \exp\{-z\};$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{U}_q^{(z,\varepsilon)}(\theta)\right]\right\}_+^q \leq C_{\varepsilon,q}\left[\underline{\mathsf{A}} \vee \underline{\mathsf{B}}\right]^q \exp\{-z\}.$$

$$\mathrm{V}^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(\mathsf{A}(\theta)\sqrt{\quad (1+\varepsilon)^2 z} + \mathsf{B}(\theta)\left[\quad (1+\varepsilon)^2 z\right]\right)$$

$$\mathrm{U}_q^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(\mathsf{A}(\theta)\sqrt{\quad (1+\varepsilon)^2 z} + \mathsf{B}(\theta)\left[\quad (1+\varepsilon)^2 z\right]\right)$$

# Bounds in general case: Payment for uniformity

$$\mathbb{P}\Big\{\Psi(\chi_\theta) \geq A(\theta)\sqrt{z} + B(\theta)z\Big\} \leq c\exp\{-z\},$$

$$\mathbb{E}\Big\{\Psi(\chi_\theta) - \Big[A(\theta)\sqrt{z} + B(\theta)z\Big]\Big\}_+^q \leq c_q\Big[A(\theta) \vee B(\theta)\Big]^q \exp\{-z\}$$

**Proposition 1.** $\forall \vec{s} \in \mathbb{S} \times \mathbb{S}$, $\forall \varepsilon \in \big(0, \sqrt{2} - 1\big]$, $\forall z \geq 1$

$$\mathrm{P}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{V}^{(z,\varepsilon)}(\theta)\right] \geq 0\right\} \leq C_\varepsilon \exp\{-z\};$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{U}_q^{(z,\varepsilon)}(\theta)\right]\right\}_+^q \leq C_{\varepsilon,q}\left[\underline{A} \vee \underline{B}\right]^q \exp\{-z\}.$$

$$\mathrm{V}^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(A(\theta)\sqrt{\qquad\qquad (1+\varepsilon)^2 z} + B(\theta)\left[\qquad\qquad (1+\varepsilon)^2 z\right]\right)$$

$$\mathrm{U}_q^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(A(\theta)\sqrt{\qquad\qquad (1+\varepsilon)^2 z} + B(\theta)\left[\qquad\qquad (1+\varepsilon)^2 z\right]\right)$$

# Bounds in general case: Payment for uniformity

$$\mathbb{P}\Big\{\Psi(\chi_\theta) \geq A(\theta)\sqrt{z} + B(\theta)z\Big\} \leq c\exp\{-z\},$$

$$\mathbb{E}\Big\{\Psi(\chi_\theta) - \Big[A(\theta)\sqrt{z} + B(\theta)z\Big]\Big\}_+^q \leq c_q\Big[A(\theta) \vee B(\theta)\Big]^q \exp\{-z\}$$

**Proposition 1.** $\forall \vec{s} \in \mathbb{S} \times \mathbb{S}$, $\forall \varepsilon \in \big(0, \sqrt{2}-1\big]$, $\forall z \geq 1$

$$\mathrm{P}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{V}^{(z,\varepsilon)}(\theta)\right] \geq 0\right\} \leq C_\varepsilon \exp\{-z\};$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{U}_q^{(z,\varepsilon)}(\theta)\right]\right\}_+^q \leq C_{\varepsilon,q}\left[\underline{A} \vee \underline{B}\right]^q \exp\{-z\}.$$

$$\mathrm{V}^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(A(\theta)\sqrt{P_\varepsilon(\theta)+(1+\varepsilon)^2 z} + B(\theta)\Big[P_\varepsilon(\theta)+(1+\varepsilon)^2 z\Big]\right)$$

$$\mathrm{U}_q^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(A(\theta)\sqrt{M_{\varepsilon,q}(\theta)+(1+\varepsilon)^2 z} + B(\theta)\Big[M_{\varepsilon,r}(\theta)+(1+\varepsilon)^2 z\Big]\right)$$

# Bounds in general case: Payment for uniformity

$$\mathbb{P}\Big\{\Psi(\chi_\theta) \geq A(\theta)\sqrt{z} + B(\theta)z\Big\} \leq c\exp\{-z\},$$

$$\mathbb{E}\Big\{\Psi(\chi_\theta) - \Big[A(\theta)\sqrt{z} + B(\theta)z\Big]\Big\}_+^q \leq c_q\Big[A(\theta) \vee B(\theta)\Big]^q \exp\{-z\}$$

**Proposition 1.** $\forall \vec{s} \in \mathbb{S} \times \mathbb{S}$, $\forall \varepsilon \in \big(0, \sqrt{2}-1\big]$, $\forall z \geq 1$

$$\mathrm{P}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{V}^{(z,\varepsilon)}(\theta)\right] \geq 0\right\} \leq C_\varepsilon \exp\{-z\};$$

$$\mathrm{E}\left\{\sup_{\theta\in\Theta}\left[\Psi\left(\chi_\theta\right) - \mathrm{U}_q^{(z,\varepsilon)}(\theta)\right]\right\}_+^q \leq C_{\varepsilon,q}\left[\underline{A} \vee \underline{B}\right]^q \exp\{-z\}.$$

$$\mathrm{V}^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(A(\theta)\sqrt{P_\varepsilon(\theta) + (1+\varepsilon)^2 z} + B(\theta)\Big[P_\varepsilon(\theta) + (1+\varepsilon)^2 z\Big]\right)$$

$$\mathrm{U}_q^{(z,\varepsilon)}(\theta) =$$

$$(1+\varepsilon)^4\left(A(\theta)\sqrt{M_{\varepsilon,q}(\theta) + (1+\varepsilon)^2 z} + B(\theta)\Big[M_{\varepsilon,r}(\theta) + (1+\varepsilon)^2 z\Big]\right)$$

# Bounds in general case: Payment for uniformity

**Payment for uniformity:** may "disappear", i.e. in some cases

$$\mathbf{P}_\varepsilon(\theta) = \text{const}, \quad \mathbf{M}_{\varepsilon,\mathbf{q}}(\theta) = \text{const}, \quad \theta \in \Theta_{\min}$$

$$\Theta_{\min} = \{\theta \in \Theta : \; \mathbf{A}(\theta) = \underline{\mathbf{A}}, \quad \mathbf{B}(\theta) = \underline{\mathbf{B}}\}$$

$$\mathcal{A}_\varepsilon(\theta) = (1+\varepsilon)\big[\mathbf{A}(\theta)/\underline{\mathbf{A}}\big], \qquad \mathcal{B}_\varepsilon(\theta) = (1+\varepsilon)\big[\mathbf{B}(\theta)/\underline{\mathbf{B}}\big]$$

$$\mathbf{P}_\varepsilon(\theta) = 2\big[1+\varepsilon^{-1}\big]^2 \mathcal{E}\big(\mathcal{A}_\varepsilon(\theta), \mathcal{B}_\varepsilon(\theta)\big) + \big[\ell\big(\mathcal{A}_\varepsilon(\theta)\big) + \ell\big(\mathcal{B}_\varepsilon(\theta)\big)\big]$$

$$\mathbf{M}_{\varepsilon,\mathbf{q}}(\theta) = 2\big[1+\varepsilon^{-1}\big]^2 \mathcal{E}\big(\mathcal{A}_\varepsilon(\theta), \mathcal{B}_\varepsilon(\theta)\big) + (\varepsilon + \mathbf{q}) \ln\big[\mathcal{A}_\varepsilon(\theta)\mathcal{B}_\varepsilon(\theta)\big]$$

# Bounds in general case: Payment for uniformity

**Payment for uniformity** may "disappear", i.e. in some cases

$$\mathbf{P}_{\varepsilon}(\theta) = \text{const}, \quad \mathbf{M}_{\varepsilon,\mathbf{q}}(\theta) = \text{const}, \quad \theta \in \mathbf{\Theta}_{\mathbf{min}}$$

$$\mathbf{\Theta}_{\mathbf{min}} = \{\theta \in \mathbf{\Theta} : \; \mathbf{A}(\theta) = \underline{\mathbf{A}}, \; \; \mathbf{B}(\theta) = \underline{\mathbf{B}}\}$$

**This is used in:**

- very sophisticated criteria of optimality of adaptive procedures, related to the **geometry** of the parameter set $\mathbf{\Theta}$;
- the application of **pointwise** adaptive procedures in **global** estimation e.g. estimation of functions possessing inhomogeneous smoothness (Nikolski, Besov classes).

Part II

# Probabilistic studies of statistical objects

## Non-asymptotical point of view

Let $\mathbf{X}$ be $\mathbf{d}$-dimensional random vector defined on $(\boldsymbol{\Omega}, \mathfrak{A}, \mathrm{P})$ and let $\mathbf{X_i}, \mathbf{i} = \overline{\mathbf{1}, \mathbf{n}}$ be independent copies of $\mathbf{X}$.

Let $\mathbf{W}$ be a given set of **bounded** functions $\mathbf{w} : \mathbb{R}^{\mathbf{d}} \to \mathbb{R}$.

$$\xi_{\mathbf{w}}(\mathbf{t}) = \frac{1}{\mathbf{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} \big[\mathbf{w}\big(\mathbf{X_i} - \mathbf{t}\big) - \mathbb{E}\mathbf{w}(\mathbf{X} - \mathbf{t})\big], \quad \mathbf{w} \in \mathbf{W}, \ \mathbf{t} \in \mathbb{R}^{\mathbf{d}}.$$

We will be interested in funding an upper function for

$$\|\xi_{\mathbf{w}}\|_{\infty} := \sup_{\mathbf{t} \in \mathbb{R}^{\mathbf{d}}} \big|\xi_{\mathbf{w}}(\mathbf{t})\big|, \quad \mathbf{w} \in \mathbf{W}.$$

**Remark:** The idea is to exploit the fact that Assumption 1 is verified on $\boldsymbol{\Theta} = \mathbf{W} \times \mathbb{R}^{\mathbf{d}}$ with $\boldsymbol{\Psi}(\cdot) = |\cdot|$, i.e. for

$$\big|\chi_\theta\big| = \big|\xi_{\mathbf{w}}(\mathbf{t})\big|, \quad \theta = (\mathbf{w}, \mathbf{t}).$$

# Bounds for empirical processes. Special cases

$$\mathcal{K} = \{K : \mathbb{R}^d \to \mathbb{R}\}, \quad \mathcal{H} = \otimes_{i=1}^d [h_i^{min}, h_i^{max}], \quad h^{min}, h^{max} \in \mathbb{R}^d$$

$$K_h(\cdot) = V_h^{-1} K(\cdot/h) \qquad V_h := \prod_{i=1}^d h_i$$

$$W_{\mathcal{K},\mathcal{H}} = \{w = K_h, \ (K,h) \in \mathcal{K} \times \mathcal{H}\}$$

$$W^{\otimes}_{\mathcal{K},\mathcal{H}} = \{w = K_h \star Q_{\mathfrak{h}}, \ \ (K,h),(Q,\mathfrak{h}) \in \mathcal{K} \times \mathcal{H}\}$$

**Kernel density estimation process:**

$$\xi_{K_h}(t) = \frac{1}{n}\sum_{i=1}^n \left[K_h(X_i - t) - \mathbb{E}K_h(X - t)\right]$$

**Convoluted kernel density estimation process:**

$$\xi_{K_h \star Q_{\mathfrak{h}}}(t) = \frac{1}{n}\sum_{i=1}^n \left\{\left[K_h \star Q_{\mathfrak{h}}\right](X_i - t) - \mathbb{E}\left[K_h \star Q_{\mathfrak{h}}\right](X - t)\right\}$$

## Assumption (K)

1. $\mathcal{K} \subset \mathbb{H}_d(\alpha, L)$, *for some* $\alpha > 0$, $L > 0$.
2. $\sup_{\delta\in(0,1)} \delta^{\beta}\mathfrak{E}_{\mathcal{K}} < C_\beta < \infty$, *for some* $\beta \in (0,1)$.
3. $\left|\int K\right| \geq C_{\mathcal{K}} > 0$, $\operatorname{supp}(K) \in [-1/2, 1/2]^d$, $\forall K \in \mathcal{K}$.

## Assumption (F)

$$f \in \mathbb{F} := \left\{ g : \mathbb{R}^d \to \mathbb{R} : \quad g \geq 0,\ \int g = 1,\ ||g||_\infty \leq f_\infty \right\}.$$

**Important quantity:**

$$f_{\mathcal{H}}(t) = \sup_{h\in\mathcal{H}} (V_h)^{-1} \int \mathbb{I}_{\otimes_{i=1}^{d}[-h_i,h_i]}(x-t) f(x)dx \leq f_\infty$$

$$\xi_{K_h}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[K_h(X_i - t) - \mathbb{E}K_h(X - t)\right]$$

**Theorem (Corollary of Proposition)**

$$\mathbb{P}_f\left\{\sup_{(K,h)\in\mathcal{K}\times\mathcal{H}}\left(\left\|\xi_{K_h}\right\|_\infty - \mathcal{U}_n\left[K_h\right]\right) \geq 0\right\} \leq \frac{95}{\ln(n)}$$

$$\mathcal{U}_n\left[K_h\right] = \mu\left\|f_{\mathcal{H}}\right\|_\infty ||K||_2\sqrt{\frac{\ell_n(h)}{nV_h}} + \mu^2 ||K||_\infty\left[\frac{\ell_n(h)}{nV_h}\right]$$

- $\ell_n(h) = \ln(1/V_h) \vee \ln\ln(n)$.
- The uniform bound $\mathcal{U}_n(\cdot)$ depends on the density $f$ via $\|f_{\mathcal{H}}\|_\infty$ only. Note also that $\|f_{\mathcal{H}}\|_\infty \leq \|f\|_\infty \leq f_\infty$.
- Explicit expression for $\mu = \mu(\alpha, L, \beta, C_\beta, C_{\mathcal{K}})$ is available.

**Theorem (Corollary of Proposition)**

$$\mathbb{P}_f \left\{ \sup_{(K,h)\in\mathcal{K}\times\mathcal{H}} \left( \left\|\xi_{K_h}\right\|_\infty - \mathcal{U}_n\left[K_h\right] \right) \geq 0 \right\} \leq \frac{95}{\ln(n)}$$

$$\mathcal{U}_n\left[K_h\right] = \mu \left\|f_{\mathcal{H}}\right\|_\infty ||K||_2 \sqrt{\frac{\ell_n(h)}{nV_h}} + \mu^2 ||K||_\infty \left[\frac{\ell_n(h)}{nV_h}\right]$$

**Theorem [Einmahl and Mason (2005)]**

Let $K$ be given and let $nh^d_{min} = \mathcal{O}\left(\ln n\right)$. Then

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}} \frac{\sqrt{nh^d}\left\|\xi_{K_h}\right\|_\infty}{\sqrt{\ln(1/h) \vee \ln\ln(n)}} < \infty \text{ a.s.}$$

**Key words**: uniform in bandwidth consistency, **LL** for kernel estimators.

**Theorem (Corollary of Proposition)**

$$\mathbb{P}_f\left\{\sup_{(K,h)\in\mathcal{K}\times\mathcal{H}}\left(\left\|\xi_{K_h}\right\|_\infty-\mathcal{U}_n\left[K_h\right]\right)\geq 0\right\}\leq\frac{95}{\ln(n)}$$

$$\mathcal{U}_n\left[K_h\right]=\tilde{\mu}\, f_\infty\, L\sqrt{\frac{\ln\left(1/V_h\right)\vee\ln\ln(n)}{nV_h}}$$

**Theorem [Einmahl and Mason (2005)]**

Let $K$ be given and let $nh^d_{min}=\mathcal{O}\big(\ln n\big)$. Then

$$\limsup_{n\to\infty}\sup_{h\in\mathcal{H}}\frac{\sqrt{nh^d}\left\|\xi_{K_h}\right\|_\infty}{\sqrt{\ln(1/h)\vee\ln\ln(n)}}<\infty\ \text{a.s.}$$

**Theorem (Corollary of Proposition) $nV_{h_{min}} = \mathcal{O}(\ln n)$**

$$\mathbb{P}_f \left\{ \sup_{(K,h)\in\mathcal{K}\times\mathcal{H}} \frac{\sqrt{nV_h}\,\|\xi_{K_h}\|_\infty}{\sqrt{\ln(1/V_h) \vee \ln\ln(n)}} \geq \tilde{\mu}\, f_\infty \right\} \leq \frac{95}{\ln(n)}$$

Explicit expression for $\tilde{\mu} = \tilde{\mu}(\alpha, L, \beta, C_\beta, C_{\mathcal{K}})$ is available.

**Theorem [Einmahl and Mason (2005)]**

Let $K$ be given and let $nh^d_{min} = \mathcal{O}(\ln n)$. Then

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}} \frac{\sqrt{nh^d}\,\|\xi_{K_h}\|_\infty}{\sqrt{\ln(1/h) \vee \ln\ln(n)}} < \infty \text{ a.s.}$$

$$\xi_{K_h}(t) = \frac{1}{n}\sum_{i=1}^{n}\left[K_h(X_i - t) - \mathbb{E}K_h(X - t)\right]$$

**Remark:** $\ell_n(h) \leq \ln n \quad \Rightarrow$

$$U_n\big(K_h\big) = \|f_{\mathcal{H}}\|_\infty\, \tau\, ||K||_2\sqrt{\frac{\ln(n)}{nV_h}} + \tau^2\, ||K||_\infty \left[\frac{\ln(n)}{nV_h}\right]$$

$$f_{\mathcal{H}}(t) = \sup_{h\in\mathcal{H}} (V_h)^{-1}\int \mathbb{I}_{\otimes_{i=1}^{d}[-h_i,h_i]}(x-t)f(x)dx$$

**Theorem (Corollary of Proposition)**

$$\mathbb{E}_f\left\{\sup_{W_{\mathcal{K},\mathcal{H}}}\left[\left\|\xi_{K_h}\right\|_\infty - U_n\big(K_h\big)\right]\right\}_+^q \leq c\left[\frac{f_\infty}{\sqrt{n}}\right]^q$$

$$\xi_{K_h \star Q_{\mathfrak{h}}}(t) = \frac{1}{n}\sum_{i=1}^{n}\Big\{\big[K_h \star Q_{\mathfrak{h}}\big](X_i - t) - \mathbb{E}\big[K_h \star Q_{\mathfrak{h}}\big](X - t)\Big\}$$

$$U_n^{\otimes}(K_h \star Q_{\mathfrak{h}}) = \|f_{\mathcal{H}}\|_\infty \, \gamma_{K,Q}\sqrt{\frac{\ln(n)}{nV_{h\vee\mathfrak{h}}}} + \gamma_{K,Q}^2\left[\frac{\ln(n)}{nV_{h\vee\mathfrak{h}}}\right]$$

**Theorem (Corollary of Proposition)**

$$\mathbb{E}_f\left\{\sup_{W^{\otimes}_{\mathcal{K},\mathcal{H}}}\Big[\big\|\xi_{K_h\star Q_{\mathfrak{h}}}\big\|_\infty - U_n^{\otimes}(K_h \star Q_{\mathfrak{h}})\Big]\right\}_+^q \le \bar{c}\left[\frac{f_\infty}{\sqrt{n}}\right]^q$$

- $\gamma_{K,Q} = \bar{\tau}\ ||K||_\infty ||Q||_\infty$.
- The explicit expressions for $\bar{\tau}$ and $\bar{c}$ are available.

Part III

# Sup-norm oracle inequality in density estimation

(Joint work with A. Goldenshluger)

# Density estimation.

$\mathbb{P}$ is a probability law on Borel $\sigma$-algebra of $\mathbb{R}^d$ possessing the density $f$ with respect to the Lebesgue measure.

$X^{(n)} = (X_1, \ldots, X_n), n \in \mathbb{N}^*$, where $X_i$, $i \in \mathbb{N}^*$, are i.i.d. random vectors distributed on $\mathbb{R}^d$ in accordance to $\mathbb{P}$.

**Density model**

$$X^{(n)} \sim p_n(x) = \prod_{i=1}^{n} f(x_i), \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d$$

**Goal:** to estimate the density f from the observation $X^{(n)}$.

**Objective: to propose a data-driven selection rule from the family of kernel estimators**

$$\mathcal{F}_{\mathcal{K},\mathcal{H}} = \left\{ \hat{f}_{K_h}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - \cdot), \quad K \in \mathcal{K}, \; h \in \mathcal{H} \right\}$$

With any $(K, h), (Q, \mathfrak{h}) \in \mathcal{K} \times \mathcal{H}$ we associate the estimator

$$\hat{f}_{K_h \star Q_{\mathfrak{h}}}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \left[K_h \star Q_{\mathfrak{h}}\right](X_i - \cdot)$$

and consider the following selection rule.

# Sup-norm oracle inequality in density estimation.

**Selection rule**

$$\hat{\Delta}(K,h) = \sup_{(Q,\mathfrak{h})\in\mathcal{K}\times\mathcal{H}} \Big[ \big\|\hat{f}_{K_h\star Q_{\mathfrak{h}}} - \hat{f}_{Q_{\mathfrak{h}}}\big\|_\infty - \nu\, \hat{U}\big(Q,\mathfrak{h}\big)\Big]$$

$$(\hat{K},\hat{h}) = \arg\inf_{(K,h)\in\mathcal{K}\times\mathcal{H}} \hat{\Delta}(K,h) + \nu\, \hat{U}\big(K,h\big)$$

$$\hat{U}\big(M,\eta\big) = ||\tilde{f}_{\mathcal{H}}||_\infty\, \tau\, ||M||_\infty \sqrt{\frac{\ln(n)}{nV_\eta}} + \tau^2\, ||M||^2_\infty \left[\frac{\ln(n)}{nV_\eta}\right]$$

- $\tilde{f}_{\mathcal{H}}$ is the pilot estimator of $f_{\mathcal{H}}$:

$$\tilde{f}_{\mathcal{H}}(t) = \sup_{h\in\mathcal{H}} (nV_h)^{-1} \sum_{i=1}^d \mathbb{I}_{\otimes_{j=1}^d[-h_j,h_j]}(X_i - t)$$

- The constants $\nu$ and $\tau$ (whose explicit expressions are available) are completely determined by the constants from Assumption **K**, **q** and **d**.

# Sup-norm oracle inequality in density estimation.

**Selection rule**

$$\Delta(K,h) = \sup_{(Q,\mathfrak{h})\in\mathcal{K}\times\mathcal{H}} \Big[\big\|\hat{f}_{K_h\star Q_\mathfrak{h}} - \hat{f}_{Q_\mathfrak{h}}\big\|_\infty - \nu\,\hat{U}\big(Q,\mathfrak{h}\big)\Big]$$

$$(\hat{K},\hat{h}) = \arg\inf_{(K,h)\in\mathcal{K}\times\mathcal{H}} \Delta(K,h) + \nu\,\hat{U}\big(K,h\big)$$

**Theorem (Oracle inequality)**

$$\mathbb{E}_f\Big[\big\|\hat{f}_{\hat{K},\hat{h}} - f\big\|_\infty\Big]^q \le c_1 \inf_{\mathcal{K}\times\mathcal{H}} \mathbb{E}_f\Big[\big\|\hat{f}_{K,h} - f\big\|_\infty\Big]^q + c_2\left[\frac{f_\infty}{\sqrt{n}}\right]^q$$

- The theorem is proved under Assumptions **K**, **F** and

$$nV_{min} \ge 1, \qquad \big|h_{max}\big| < \infty.$$

- The explicit expressions of $c_1$ and $c_2$ are available.

Let $\{\mathbb{H}(\vec{\gamma}, \mathbf{L}),\ \vec{\gamma} \in (\mathbf{0}, \mathbf{b}]^{\mathbf{d}}, \mathbf{L} > \mathbf{0}\}$ be the collection of anisotropic Hölder classes on $\mathbb{R}^d$. Here $\mathbf{b} > \mathbf{0}$ is an arbitrary but a priori chosen integer.

Let $\mathbf{M} : \mathbb{R}^{\mathbf{d}} \longrightarrow \mathbb{R}$ be a given compactly supported lipschitz continuous function such that $\int \mathbf{M} = 1$ and

$$\int \mathbf{M}(\mathbf{x}) \prod_{j=1}^{d} \mathbf{x}_j^{p_j} d\mathbf{x}_j = \mathbf{0},\ \forall \vec{\mathbf{p}} \in \mathbb{N}^{\mathbf{d}} :\ \mathbf{1} \le \mathbf{p}_1 + \cdots + \mathbf{p}_d \le \mathbf{b}$$

Let $\hat{\mathbf{f}}_{\mathbf{M},\hat{\mathbf{h}}}$ be the estimator chosen in accordance with the proposed selection rule, where $\mathcal{K} = \{\mathbf{M}\}$.

**Theorem (Adaptation).** $\forall\ \vec{\gamma} \in (0, b]^d,\ L > 0,\ q > 0$

$$\sup_{f \in \mathbb{H}(\vec{\gamma}, L)} \mathbb{E}_f \left[ \left\| \hat{\mathbf{f}}_{\mathbf{M},\hat{\mathbf{h}}} - \mathbf{f} \right\|_\infty \right]^q \asymp \mathbf{L}^{\frac{2q}{2\gamma+1}} \left( \frac{\ln n}{n} \right)^{\frac{2q\gamma}{2\gamma+1}}, \quad \frac{1}{\gamma} = \sum_{j=1}^{d} \frac{1}{\gamma_j}.$$