Adaptation via L_p -norm in the regression model.

NGUYEN Ngoc Bien

Laboratoire d'Analyse, Topologie et Probabilités
Université de Provence

CIRM, April 16, 2012

Outline

1. Introduction
   - Model
   - Conditions
   - Collection of estimators and the selection rule
2. Results
3. Adaptation
   - anisotropic Hölder classes
   - Adaptation

Let $Y_i = f(X_i) + \xi_i \quad i = \overline{1, n}$ where:

1. (X_i, Y_i)-observation.
2. (X_i) are i.i.d and uniformly distributed on $[0, 1]^d$.
3. (ξ_i)'s are i.i.d and independent of (X_i).
4. $f \in \mathcal{F} := \{g : [0, 1]^d \to \mathbb{R} \, , \ \|g\|_\infty \leqslant f_\infty\}$

Our goal: Estimate the function f from the observation $(X_i, Y_i), i = \overline{1, n}$.

Risk:

$$\mathcal{R}_s^q(\hat{f}, f) := \left(\mathbb{E}_f \|\hat{f} - f\|_{s,\nu}^q\right)^{1/q}$$

$$\mathcal{R}_s^q(\hat{f}, \mathbf{F}) := \sup_{f \in \mathbf{F}} \left(\mathbb{E}_f \|\hat{f} - f\|_{s,\nu}^q\right)^{1/q}$$

Where the mesure $d\nu = 1_{[\delta, 1-\delta]^d}(x)dx$ with $0 < \delta < 1/2$ given.

We need the assumption on the noise,

Assumption N

1. There exists $c > 0, \alpha > 0$ such that $\mathbb{P}\{|\xi_1| \geqslant x\} \leqslant c\exp\{-x^{\alpha}\} \ \ \forall x > 0.$
2. There exists $p > 1$ and $P > 0$ such that $\mathbb{E}|\xi|^p \leqslant P$

Estimator linear.

$$\hat{f}_h(t) := \frac{1}{n}\sum_{i=1}^{n} K_h(X_i - t)Y_i \quad , h \in H$$

where

- $h := (h_1, ..., h_d)$-bandwidth, $K : \mathbb{R}^d \to \mathbb{R}$-kernel, $K_h(\cdot) := V_h^{-1}K(\cdot/h)$, $V_h := h_1 \cdots h_d$ and if $x, h \in \mathbb{R}^d$ then $x/h := (x_1/h_1, ..., x_d/h_d)$
- $H := \left\{h \in [h_{min}, 1]^d : V_h \leqslant V_{max}\right\}$ where $0 < h_{min} < 1$ and $V_{max} > 0$.

Assumptions .

We need the assumption on the kernel K

Assumption K

1. $\mathrm{supp} K \subset [-1/2, 1/2]^d$
2. There exists $k > 0$ such that $\forall x, y \in \mathbb{R}^d$, we have $|K(x) - K(y)| \leqslant k\|x - y\|$ where $\|\cdot\|$ is the euclidean norm.
3. $\int K(x)dx = 1$

Putting

$$\hat{f}_{h,\eta}(t) := \frac{1}{n}\sum_{i=1}^{n} (K_h * K_\eta - K_h)(X_i - t)Y_i \ , h, \eta \in H$$

where $*$ is the convolution.

Selection rule

- $$\hat{R}_h := \sup_{\eta \in H} \left\{ \|\hat{f}_{\eta,h} - \hat{f}_\eta\|_{s,\nu} - C\frac{1}{\sqrt{nV_h}} \right\}_+ + C\frac{1}{\sqrt{nV_h}}$$
- $\hat{h} := \arg\inf_{h\in H} \hat{R}_h$
- $\hat{f} := \hat{f}_{\hat{h}}$

where $C := C(k, d, s, q, \alpha, f_\infty)$ if N1, and
$C := C(k, d, s, q, p, P, f_\infty)$ if N2

Selection rule

Remark

Because H is a compact, $\hat{f}_h(\cdot)$ is continous a.s then $\hat{h}$ is measurable.

Results.

Theorem 1

If the assumption N1 holds, then we have:

$$\mathcal{R}_s^q(\hat{f}, f) \leqslant (1+2k) \inf_{h \in H} \left\{ \mathcal{R}_s^q(\hat{f}, f) + C_1 \frac{1}{\sqrt{nV_h}} \right\}$$

$$+ C_2 n \ln^{3d}(h_{min}^{-1}) \exp\left\{ -\frac{1}{4q} V_{max}^{-\frac{\alpha}{s(\alpha+1)}} \right\} \quad , \forall f \in \mathcal{F}$$

where C_1, C_2 depend only on $f_\infty, k, s, \alpha, d$ and q

Results.

Theorem 2

If the assumption N2 holds, then we have:

$$\mathcal{R}_s^q(\hat{f}, f) \leqslant (1+2k) \inf_{h \in H} \left\{ \mathcal{R}_s^q(\hat{f}, f) + C_3 \frac{1}{\sqrt{nV_h}} \right\}$$

$$+ C_4 n \ln^{3d}(h_{min}^{-1}) V_{max}^{p/(3qs)} \quad , \forall f \in \mathcal{F}$$

where C_3, C_4 depend only on f_∞, k, s, p, P, d and q

Definition

Let $\beta = (\beta_1, ..., \beta_d)$, $\beta_i > 0$ and $L > 0$. We say that the function $f : \mathbb{R}^d \to \mathbb{R}$ belongs to the anisotropic Holder class $H_d(\beta, L)$ of function if:

For all $i = 1, ..., d$ and all $t \in \mathbb{R}$

$$\sup_{x_1,...,x_d \in \mathbb{R}^d} \left| D_i^{\lfloor \beta_i \rfloor} f(x_1, ..., x_i + t, ..., x_d) - D_i^{\lfloor \beta_i \rfloor} f(x_1, ..., x_i, ..., x_d) \right|$$

$$\leq L \left| t \right|^{\beta_i - \lfloor \beta_i \rfloor}$$

Here $D_i^k f$ denotes the kth order partial derivative of f with respect to the variable t_i and $\lfloor t \rfloor$ is the largest integer strictly less than t.

Adaptation.

We define also $\phi_n(\beta) = n^{-\bar{\beta}/(2\bar{\beta}+1)}$ where $1/\bar{\beta} = \sum_{i=1}^{d} 1/\beta_i$
$\mathfrak{H} := \left\{ H_d(\beta, L) : 0 < \beta_i < l, i = \overline{1, d}, L > 0 \right\}$ where $l > 0$ fixed.

Additional assumption on K

$\int_{\mathbb{R}^d} K(t) t^k dt = 0 \quad \forall\, |k| = 1, ..., \lfloor l \rfloor - 1$ where $k = (k_1, ..., k_d)$ is multi-index, $|k| = k_1 + \cdots + k_d \quad t^k = t_1^{k_1} \cdots t_d^{k_d}$ for $t = (t_1, ..., t_d)$

We set $h_{min} = 1/n$ and $V_{max} = n^{-d/(2l+d)}$ then we have

Theorem 3

For all $s \geqslant 1$, $H_d(\beta, L) \in \mathfrak{H}$, assume that $p \geqslant 9qs(l + 1/2)$ if N2, then

$$\lim_{n\to\infty} \sup \left[\phi_n^{-1}(\beta)\mathcal{R}_s^q(\hat{f}, H_d(\beta, L))\right] < +\infty$$

Remark

It is well-known that $\phi_n(\beta)$ is the rate-minimax over the space function $H_d(\beta, L)$. Then our theorem precedent indice the adaptation of estimator $\hat{f}$ over the class $\mathfrak{H}$

Proof of theorem 1 and 2

$$
\begin{aligned}
\|\hat{f} - f\|_s &\leq \|\hat{f}_{\hat{h}} - \hat{f}_{\hat{h},h}\|_s + \|\hat{f}_{\hat{h},h} - \hat{f}_h\|_s + \|\hat{f}_h - f\|_s \\
&\leq \|\hat{f}_h - f\|_s + \left[\|\hat{f}_{\hat{h}} - \hat{f}_{\hat{h},h}\|_s - C\frac{1}{\sqrt{nV_{\hat{h}}}}\right]_+ + C\frac{1}{nV_{\hat{h}}} \\
&+ \left[\|\hat{f}_{h,\hat{h}} - \hat{f}_h\|_s - C\frac{1}{\sqrt{nV_h}}\right]_+ + C\frac{1}{\sqrt{nV_h}} \\
&\leq \|\hat{f} - f\|_s + \left\{\sup_{\eta\in H}\left[\|\hat{f}_\eta - \hat{f}_{\eta,h}\|_s - C\frac{1}{\sqrt{nV_\eta}}\right]_+ + C\frac{1}{\sqrt{nV_h}}\right\} \\
&+ \left\{\sup_{\eta\in H}\left[\|\hat{f}_\eta - \hat{f}_{\eta,\hat{h}}\|_s - C\frac{1}{\sqrt{nV_\eta}}\right]_+ + C\frac{1}{\sqrt{nV_{\hat{h}}}}\right\} \\
&= \|\hat{f}_h - f\|_s + \hat{R}_h + \hat{R}_{\hat{h}} \\
&\leq \|\hat{f}_h - f\|_s + 2\hat{R}_h
\end{aligned}
$$

Proof of theorem 1 and 2

To bound $\hat{R}_h$ we have to bound

$$M_{s,h}(f) := \left(\mathbb{E}_f \sup_{\eta \in H} \left[\|\hat{f}_\eta - \hat{f}_{\eta,h}\|_s - C \frac{1}{\sqrt{nV_\eta}} \right]_+^q \right)^{1/q}$$

Writing
$\hat{f}_{\eta,h}(t) - \hat{f}_\eta(t) =$bias $+$ stochastic error$:= A_{h,\eta}(t) + B_{h,\eta}(t)$.
The hardest work is to bound

$$M_h^{(1)}(f) := \left(\mathbb{E}_f \sup_{\eta \in H} \left[\|B_{h,\eta}\|_s - C^{(1)} \frac{1}{\sqrt{nV_\eta}} \right]_+^q \right)^{1/q}$$

The main technical tools used in our derivations are uniform bounds on $\mathbb{L}_p$-norms of empirical processes developed by Goldensluger and Lepski [2010].

Refercenes.

- A.Goldensluger and O.Lepski : *Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality* To appear in Ann. Stat.
- A.Goldensluger and O.Lepski : *Structural adaptation via $\mathbb{L}_p$-norm oracle inequalities*, Probab.Theory Ralat. Fields 143,41-71.
- A.Goldensluger and O.Lepski : *Uniform bounds for norms of independent random functions*, Ann. Probab 39, 2318-2384.
- A.Goldensluger and O.Lepski : *Universal estimation routines in non parametric statistics*, Manuscrit.
- G. Kerkyacharian, O. Lepski and D. Picard :*Nonlinear estimation in anisotropic multi-index denoising*, Probab. Theory Relat. Fields 121, 137-170.
- O.Lepski and B.Y.Levit : *Universal pointwise selection rule in multivariate function estimation*, Bernoulli 14, 1150-1190.