About the Goldenshluger-Lepski methodology for bandwidth selection

Claire Lacour[1], Pascal Massart[1], Vincent Rivoirard[2]

1-Université Paris-Sud, 2-Université Paris Dauphine

Statistical framework

$X_1, \ldots, X_n$ i.i.d. real random variables with unknown density f
$\hat{f}_h$ classical kernel estimator

$$\hat{f}_h(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - X_i)$$

with $K_h(\cdot) = \frac{1}{h} K\left(\frac{\cdot}{h}\right)$ and K a given kernel

Quadratic loss $\mathbb{E}\|\hat{f}_h - f\|^2$ where $\|.\|$ is the L^2 norm

Bias estimation

Notation $f_h := \mathbb{E}[\hat{f}_h] = K_h * f$

Bias-variance decomposition:

$$\mathbb{E}\|\hat{f}_h - f\|^2 = \|f_h - f\|^2 + \mathbb{E}\|f_h - \hat{f}_h\|^2 \approx \underbrace{\|f_h - f\|^2}_{B^2(h)} + \underbrace{\frac{\|K_h\|^2}{n}}_{V(h)}$$

Idea: estimator $\hat{B}^2(h)$ of $B^2(h)$ and then

$$\hat{h} = \operatorname*{argmin}_{h \in \mathcal{H}} \{\hat{B}^2(h) + V(h)\}$$

Some heuristics

With high probability

$$\|\hat{f}_h - \hat{f}_{h'}\|^2 \approx \|f_h - f_{h'}\|^2 + \frac{\|K_h - K_{h'}\|^2}{n}$$

$\hookrightarrow \hat{B}^2(h) = \sup_{h' \leq h}\{\|\hat{f}_h - \hat{f}_{h'}\|^2 - \frac{\|K_h - K_{h'}\|^2}{n}\}$

$$\|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 \approx \|f_h - f_{h_{\min}}\|^2 + \frac{\|K_h - K_{h_{\min}}\|^2}{n}$$

$\hookrightarrow \hat{B}^2(h) = \{\|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 - \frac{\|K_h - K_{h_{\min}}\|^2}{n}\}$

Goldenshluger-Lepski method

$\mathcal{H} \in \mathbb{R}_+^*$ finite subset of bandwiths

$$\begin{cases} \hat{B}^2(h) = \sup_{h' \leq h} \left[\|\hat{f}_h - \hat{f}_{h'}\|_2 - \text{pen}(h') \right]_+ \\ \hat{h} = \underset{h \in \mathcal{H}}{\text{argmin}} \ \{\hat{B}^2(h) + \text{pen}(h)\} \end{cases}$$

Actually:

- more general
- $\|\hat{f}_h - \hat{f}_{h'}\|_2 \longrightarrow \|\hat{f}_{h,h'} - \hat{f}_{h'}\|_2$ with $\hat{f}_{h,h'}$ auxiliary estimators (not important here)

Here Penalty="Majorant"= a Variance $= a\dfrac{\|K\|^2}{nh}$

Oracle inequality

$\tilde{B}(h) := \max(\sup_{h' \leq h} \|f_{h'} - f_h\|, \|f - f_h\|) \approx$ bias

Theorem

Assume that $\|f\|_\infty < \infty$ *and* $K \geq 0$ *unimodal with mode 0, and* $\mathcal{H} \subset [n^{-1}, \log^{-2}(n)]$. *If* $\mathrm{pen}(h) = a\|K\|_2^2/(nh)$ *with* $a > 1$, *then*

$$\mathbb{E}\|\hat{f}_{\hat{h}} - f\|^2 \leq C_0(a) \min_{h \in \mathcal{H}} \left\{ \tilde{B}^2(h) + a\frac{\|K_h\|^2}{n} \right\} + o(n^{-1})$$

Ccl: the method works well if $a > 1$

But what if a small? And how to choose a in practice?

Minimal penalty

Theorem (L., Massart, 2016)

Assume that $\|f\|_\infty < \infty$ *and* K *good chosen, and choose* $\mathcal{H} = \{e^{-k}, \lceil 2\log\log n\rceil \le k \le \lfloor \log n \rfloor\}$. *If* $\mathrm{pen}(h) = a\|K\|_2^2/(nh)$ *with* $a < 1$, *then* $\exists C(f,a,K) > 0$ *s.t., for* n *large enough,*

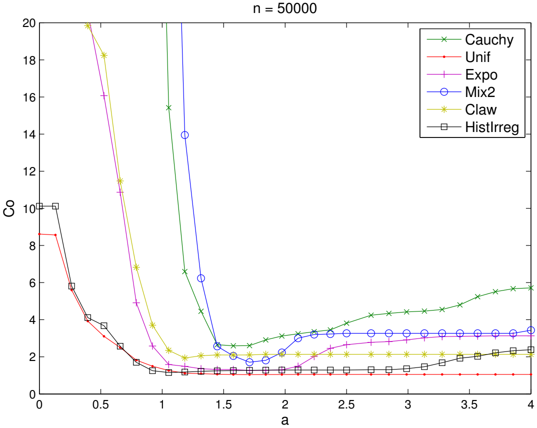
$$\mathbb{P}(\hat{h} \ge 3h_{\min}) \le C(\log n)^2 \exp(-(\log n)^2/C)$$

i.e. $\hat{h} < 3h_{\min}$ *with high probability. Consequently*

$$\liminf_{n\to\infty} \mathbb{E}\|\hat{f}_{\hat{h}} - f\|^2 > 0$$

Ccl: the method fails if $a < 1$, risk explosion

Simulations



Oracle constant C_0 as a function of a, for 6 examples of density, where $C_0 = \tilde{\mathbb{E}} \frac{\|\hat{f}_{\hat{h}} - f\|^2}{\min_{h \in \mathcal{H}} \|\hat{f}_h - f\|^2}$

Issue of calibration

- visible explosion, and a_{opt} very close to the jump
- jump not always at $a = 1$

Not possible to choose $a = 1$ in practice
$\longrightarrow$ best idea: to detect the jump $\hat{a}_J$, and then $\hat{a} = 1.1\hat{a}_J$
but not comfortable : optimal to close to minimal...

Another method to separate optimal penalty from minimal penalty:

$$\begin{cases} \hat{B}^2(h) = \sup_{h' \leq h} \left[\|\hat{f}_h - \hat{f}_{h'}\|^2 - \text{pen}_1(h, h') \right]_+ \\ \hat{h} = \underset{h \in \mathcal{H}}{\text{argmin}} \, \{\hat{B}^2(h) + \text{pen}_2(h)\} \end{cases}$$

A degenerate case of the GL method

Theorem (L., Massart, Rivoirard, *work in progress*)

If $\mathrm{pen}_1(h,h') = a\|K_h - K_{h'}\|^2/n$ *and* $\mathrm{pen}_2(h) = b\|K_h\|^2/n$ *then*

- *if* $a > 1$ *and* $b > 0$*: oracle inequality*
- *if* $0 \leq a < 1$ *and* $b < b_{crit}(a,K)$*:* $\hat{h} \approx h_{\min}$
- *if* $0 \leq a < 1$ *and* $b > b_{crit}(a,K)$*: oracle inequality*

Csq: we can choose a small... and even $a = 0$!

$\hookrightarrow$ new method

$$\begin{cases} \hat{B}^2(h) = \sup_{h' \leq h} \left[\|\hat{f}_h - \hat{f}_{h'}\|^2 \right] \approx \|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 \\ \hat{h} = \underset{h \in \mathcal{H}}{\operatorname{argmin}} \left\{ \hat{B}^2(h) + \mathrm{pen}_2(h) \right\} \end{cases}$$

A new method for bandwidth selection (1/2)

$$\hat{h} = \underset{h \in \mathcal{H}}{\text{argmin}} \left\{ \|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 + \text{pen}(h) \right\}$$

Heuristic 1:

$$\hat{f}_{h_{\min}}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_{\min}}(x - X_i) \underset{h_{\min}\to 0}{\longrightarrow} \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(x)$$

$$\Rightarrow \langle \hat{f}_h, \hat{f}_{h_{\min}} \rangle \underset{h_{\min}\to 0}{\longrightarrow} \frac{1}{n}\sum_{i=1}^{n} \hat{f}_h(X_i)$$

$$\hat{h} \approx \underset{h \in \mathcal{H}}{\text{argmin}} \left\{ \|\hat{f}_h\|^2 - \tfrac{2}{n}\sum_{i=1}^{n} \hat{f}_h(X_i) + \|\hat{f}_{h_{\min}}\|^2 + \text{pen}(h) \right\}$$

penalized least-squares contrast

method of Lerasle-Magalhães-Reynaud (2015)

Link with regression: $\hat{h} = \underset{h \in \mathcal{H}}{\text{argmin}} \left\{ \|\hat{f}_h - Y\|_n^2 + \text{pen}(h) \right\}$

A new method for bandwidth selection (2/2)

$$\hat{h} = \underset{h\in\mathcal{H}}{\text{argmin}} \{\|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 + \text{pen}(h)\}$$

Heuristic 2:

$$\begin{aligned} B^2(h) &\approx \|f_h - f_{h_{\min}}\|^2 \approx \|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 - \frac{\|K_h - K_{h_{\min}}\|^2}{n} \\ &\approx \|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 - \frac{\|K_h\|^2}{n} + 2\frac{\langle K_h, K_{h_{\min}}\rangle}{n} - \frac{\|K_{h_{\min}}\|^2}{n} \end{aligned}$$

To minimize $\{B^2(h) + b\frac{\|K_h\|^2}{n}\}$ is equivalent to minimize

$$\|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 + \underbrace{2\frac{\langle K_h, K_{h_{\min}}\rangle}{n} + (b-1)\frac{\|K_h\|^2}{n}}_{\text{pen}(h)}$$

Minimal penalty

$$\hat{h} = \underset{h \in \mathcal{H}}{\text{argmin}} \left\{ \|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 + 2\frac{\langle K_h, K_{h_{\min}} \rangle}{n} + (b-1)\frac{\|K_h\|^2}{n} \right\}$$

Theorem 1 (L., Massart, Rivoirard, 2016)

Assume that $\|f\|_\infty < \infty$ *and* $\|K\|_\infty \|K\|_1 n^{-1} \leq h_{\min} \ll \log^{-2}(n)$ *and* $\|f_{h_{\min}} - f\|^2 = o(1)$
If $b < 0$, $\forall q > 0$, *for* n *large enough,*

$$\hat{h} \leq C(b) h_{\min} \qquad \text{with probability } 1 - n^{-q}$$

and then $\liminf_{n \to \infty} \mathbb{E}\|\hat{f}_{\hat{h}} - f\|^2 > 0$

Oracle inequality

$$\hat{h} = \operatorname*{argmin}_{h \in \mathcal{H}} \left\{ \|\hat{f}_h - \hat{f}_{h_{\min}}\|^2 + 2\frac{\langle K_h, K_{h_{\min}} \rangle}{n} + (b-1)\frac{\|K_h\|^2}{n} \right\}$$

Theorem 2 (L., Massart, Rivoirard, 2016)

Assume $\|f\|_\infty < \infty$ *and* $h_{\min} \geq \|K\|_\infty \|K\|_1 / n$. *Let* $\epsilon \in (0,1)$.
If $b > 0$, $\forall x > 0$, *with probability* $1 - C_1 |\mathcal{H}| e^{-x}$

$$\|\hat{f}_{\hat{h}} - f\|^2 \leq C_0(b) \min_{h \in \mathcal{H}} \|\hat{f}_h - f\|^2 + C_2 \|f_{h_{\min}} - f\|^2 + C_3 \frac{\|f\|_\infty x^3}{n}$$

with $C_0(b) = \begin{cases} b + \epsilon & \text{if } b > 1 \\ 1 + \epsilon & \text{if } b = 1 \quad \leftarrow \textit{optimality} \\ 1/b + \epsilon & \text{if } 0 < b < 1 \end{cases}$

Conclusion

- We just prove

$$b = 0 \qquad \text{pen}_{\min} = 2\frac{\langle K_h, K_{h_{\min}}\rangle}{n} - \frac{\|K_h\|^2}{n}$$

$$b = 1 \qquad \text{pen}_{\text{opt}} = 2\frac{\langle K_h, K_{h_{\min}}\rangle}{n}$$

minimal different from the optimal: good news for calibration
Examples: $\text{pen}_{\text{opt}} = \text{pen}_{\min} * 2$ for rectangular kernel,
$\text{pen}_{\text{opt}} = \text{pen}_{\min} * \frac{2\sqrt{2}}{2\sqrt{2}-1}$ for Gaussian kernel

- Simple to implement, less comparisons than for Lepski method: numerically faster (numerical experiments in progress...)

Introduction

Calibration for the Goldenshluger-Lepski method

Method of comparison with the worse

Future works

Future works

- multivariate case
- further exploration of Goldenshluger-Lepski method
- other loss functions: Hellinger or L^1 loss (more appropriate for densities)