On bandwidth selection problems in nonparametric trend estimation under martingale difference errors

Sana Louhichi

Based on a common work with Karim Benhenni (UGA) and Didier A. Girard (LJK, CNRS) [Bernoulli 2022 and Advances in Pure and Applied Mathematics 2022]

Closing ECODEP conference & Opening new windows & PI's Emeritus conference
IHP, Amphitheater Hermite: February 12-14, 2024

Hyperparameters for Machine learning models

Unlike parameters which are learned from data, hyperparameters are defined by the data scientist and influence the learning process and the model's ability to generalize.

- Models: Regression, Knn, SVM, Neural Networks, random forests
- Examples: Learning rate, batch size, and number of epochs in neural network training; the depth of a decision tree, the bandwidth for kernel estimation...

The choice of hyperparameters has a profound impact on the performance of machine learning models. Poorly selected hyperparameters can result in models that underperform or are prone to overfitting, where they memorize the training data but fail to generalize to new data.

Hyperparameters and Regression models

The general model assumes a relationship between Y, the variable being explained and one or more variables doing the explaining X.

- The data set $(X_i, Y_i)_{1 \leq i \leq n}$ with unknown common distribution as (X, Y).
- **Model:**

$$Y_i = f(X_i) + \epsilon_i, \;\; E(\epsilon_i / X_i) = 0 \; a.s.$$

  f **Regression function of Y on X, (ϵ_i): noise.**
- Purpose: Estimate f based on the data set $(X_i, Y_i)_{1 \leq i \leq n}$.
- **Criteria: based on RSS(f) (minimising the noise**

$$RSS(f) = \sum_{i=1}^{n} (Y_i - f(X_i))^2).$$

Hyperparameters and Regression models

- Linear regression:
  $\hat{\beta} = argmin_\beta \left\{ \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \Psi(\beta) \right\}.$
  $\Psi(\beta) = \|\beta\|_2^2, \|\beta\|_1 \cdots$ .. Ridge, Lasso, Elastic net...
- Kernel methods and local regression:
  bandwidth h the width of the kernel.
- Projection estimator (basis functions and dictionary methods):
  N number of basis functions.
- Cubic smoothing spline:
  $argmin_f \left\{ \sum_{i=1}^n (y_i - f(x_i))^2 + \lambda J(f) \right\}$
  $J(f) = \int f''^2(x)dx.$
  λ the multiplier of the penalty term.

Linear estimators

For any x,

$$\hat{f}(x) = \sum_{i=1}^{n} l_i(x) Y_i, \quad \sum_{i=1}^{n} l_i(x) = 1.$$

In a matricial form, $Y = (Y_1, \cdots, Y_n)^t$

$$(\hat{f}(X_1), \cdots, \hat{f}(X_n))^t = LY$$

$L = (L_j(X_i))_{1 \leq i,j \leq n}$

$\nu = trace(L)$: the effective degree of freedom.
Bias-variance Tradeoff, overfitting, underfitting.

How to choose the hyper-parameter? Minimizer of the ASE

$$ASE(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(\hat{f}(x_i) - f(x_i))^2,$$

$$\hat{\lambda}_n = argmin_\lambda ASE(\lambda)$$

Problem ?

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}(x_i))^2$$

Problem ?
A poor estimate of MASE(λ): it is biased downwards and typically leads to overfitting. The reason is that we are using the data twice: to estimate the function and to estimate the risk.

How to choose the hyper-parameter? Cross Validation

Definition

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{f}^{-i}(x_i))^2$$

where $\hat{f}^{-i}$ is the estimator obtained by omitting the ith pair (x_i, Y_i).

$$\hat{f}(x) = \sum_{j=1}^{n} l_{j,n}(x) Y_j \quad \hat{f}^{-i}(x) = \sum_{j=1}^{n} l_{j,n}^{-i}(x) Y_j,$$

$$l_{j,n}^{-i}(x) = \frac{l_{j,n}(x)}{\sum_{j,\ j \neq i} l_{j,n}(x)} \mathbb{1}_{j \neq i}.$$

How to choose the hyper-parameter? GCV

Lemma

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \hat{f}(x_i))^2}{(1 - l_i(x_i))^2}$$

Definition

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_i - \hat{f}(x_i))^2}{(1 - \nu/n)^2}$$

$$\nu = tr(L) = \sum_{i=1}^{n} l_i(x_i)$$

$$(1-x)^{-2} = 1 + 2x + \ldots$$

How to choose the hyperparameter? Mallows criterion

Definition

$$C_p := C_p(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}(x_i))^2 + 2\frac{\nu}{n}\hat{\sigma}^2,$$

where,

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}(x_i))^2, \;\; \nu = tr(L) = \sum_{i=1}^{n} l_i(x_i).$$

For i.i.d, an unbiased estimator of MASE is, if σ^2 is known,

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}(x_i))^2 + 2\frac{\nu}{n}\sigma^2 - \sigma^2.$$

In the dependent case, an unbiased estimator of MASE, if $\sigma^2_{i,j}$ is known,

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{f}(x_i))^2 + \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} l_j(X_i)\mathrm{Cov}(\epsilon_i, \epsilon_j) - \sigma^2.$$

$$\lambda_n = argmin_h MASE(\lambda) \quad \hat{\lambda}_n = argmin_\lambda ASE(\lambda),$$
$$\hat{\lambda}_{ML} = argmin_\lambda C_p(\lambda) \quad \hat{\lambda}_G = argmin_\lambda GCV(\lambda)$$

All those windows are nearly equivalent in probability.

Paper: W. Härdle, P. Hall and J. S. Marron (1988). How far are automatically chosen regression smoothing parameters from their optimum? Journal of the American Statistical Association 83, 86-95.

Purposes

- Asymtotic behaviors: "asymptotic optimality" $\frac{\hat{\lambda}}{\hat{\lambda}_n} \to 1$, prob.
- comparison, confidence sets: asymptotic normality $v_n(\hat{\lambda} - \hat{\lambda}_n)$
- The excess of error: (comparison of risks)
  $w_n(ASE(\hat{\lambda}) - ASE(\hat{\lambda}_n))$

For dependent observations
Errors with no necessarily normal law

Criterion under dependence uses estimates of $\mathrm{Cov}(\epsilon_i, \epsilon_j)$
For $i \neq j$, $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ for MDS.

Assumptions

Assume that the errors $(\epsilon_i)_{i\geq 0}$:

- form a strictly stationary MDS with respect to some natural filtration $(\mathcal{F}_i)_{i\geq 1}$, i.e, for any $i > 0$, ϵ_i is $\mathcal{F}_i$-measurable and $\mathbb{E}(\epsilon_i|\mathcal{F}_{i-1}) = 0$ almost surely. Suppose also that $\mathbb{E}(\epsilon_1^{2p}) < \infty$ for some $p > 8$.

Examples

- With ARCH(1) errors:

$$
\begin{aligned}
&Y_i = f(x_i) + \epsilon_i \\
&\epsilon_n = e_n \sqrt{\sigma^2(1-\alpha) + \alpha\epsilon_{n-1}^2},\ 0 \le \alpha < 1,\ \sigma^2 > 0 \\
&(e_i)\ iid,\ normal\ law \\
&e_n\ indep\ (\epsilon_0, \cdots, \epsilon_{n-1})
\end{aligned}
$$

Robert F. Engle, 1982. Econometric and finance problems. $Var(\epsilon_n | \epsilon_{n-1}) = \sigma^2(1-\alpha) + \alpha\epsilon_{n-1}^2$.

Examples

- **SV with Log-normal volatility sequences:** $\epsilon_n = \sigma_n Z_n$ (Z_n) iid centered independent of $(\sigma_n)_n$. the volatility sequence $(\sigma_i)_{i\in\mathbb{N}}$ is an exponential weight of a Gaussian moving average:

$$\sigma_i = \beta \exp\left(\sum_{j=0}^{\infty} \gamma^j \eta_{i-j}\right), \ \beta > 0, \ |\gamma| < 1.$$

$(\eta_i)_i$ iid with centered normal law.

Taylor (1986): *Modelling Financial Time Series*.

How to choose h? Minimizer of the MASE

K : $[-1,1]$-compactly supported Kernel

$$f \in \mathcal{C}^2,\ x_i = \frac{i}{n},\ Y_i = f(x_i) + \epsilon_i,\ 1 \le i \le n$$

Lemma

Define,
$$D_n(h) = \frac{h^4}{4}\int_0^1 u(x) f''^2(x)dx \left(\int_{-1}^1 t^2 K(t)dt\right)^2 + \frac{\sigma^2}{nh}\left(\int_0^1 u(x)dx\right)\int_{-1}^1 K^2(y)dy.$$

Then for any $n \ge 1$ and $h \in]0,\epsilon[$,

$$MASE(h) = D_n(h) + O\left(\frac{1}{n}\right) + O(h^5) + O\left(\frac{1}{n^2h^4}\right) + \frac{O(h)}{nh}.$$

How to choose h? Minimizer of the MASE

Let $h_n^* = argmin_{h>0} D_n(h)$. Clearly, if $\int_0^1 u(x) f''^2(x) dx \neq 0$ then

$$h_n^* = n^{-1/5} \left(\frac{\left(\int_0^1 u(x)dx\right) \int_{-1}^1 K^2(y) dy \sigma^2}{\int_0^1 u(x) f''^2(x) dx \left(\int_{-1}^1 t^2 K(t) dt\right)^2} \right)^{1/5} =: cn^{-1/5}.$$

Problem ?

Our first result proves that for MDS of errors, the selected bandwidths $h_n, h_n^*, \hat{h}_n$ and $\hat{h}_{ML}$ are nearly equivalent.

Proposition 1, Benhenni, Girard, Louhichi (2022).

It holds, under the above notations and conditions,

$$\frac{h_n^*}{h_n}, \frac{\hat{h}_n}{h_n}, \frac{\hat{h}_{ML}}{h_n}$$

converge all in probability to 1 as n tends to infinity.

Our second result gives the rate at which $\hat{h}_n - \hat{h}_{ML}$ converges in distribution to a centered normal law.

Theorem 1, Benhenni, Girard, Louhichi (2022).

Suppose that the above conditions are satisfied. Suppose also, that there exists a positive decreasing function Φ defined on $\mathbb{R}^+$ satisfying

$$\sum_{s=1}^{\infty} s^4 \Phi(s) < \infty,$$

and for any positive integer q less than 6, $1 \leq i_1 \leq \cdots \leq i_k < i_{k+1} \leq i_q \leq n$ such that $i_{k+1} - i_k \geq \max_{1 \leq l \leq q-1}(i_{l+1} - i_l)$

$$|\mathrm{Cov}(\epsilon_{i_1} \cdots \epsilon_{i_k}, \epsilon_{i_{k+1}} \cdots \epsilon_{i_q})| \leq \Phi(i_{k+1} - i_k). \tag{1}$$

Then

$$n^{3/10}(\hat{h}_n - \hat{h}_{ML})$$

converges in distribution to a centered normal law with variance Σ^2 given by

$$\Sigma^2 = \frac{4\sigma^{12/5} B^{1/5}}{5A^{6/5}} \left(\int t^2 K(t)dt\right)^2 \int_0^1 u^2(x) f''^2(x)dx$$
$$+ \frac{16\sigma^{12/5}}{5A^{1/5}B^{4/5}} \int_0^1 u^2(x)dx \int_0^1 (K-G)^2(u)du,$$

where $\sigma^2 = \mathbb{E}(\epsilon_1^2)$, G is the function defined for any $x \in \mathbb{R}$ by $G(x) = -xK'(x)$ and

$$A = \int_0^1 u(x)f''^2(x)dx \int t^2 K(t)dt, \;\; B = \int_0^1 u(x)dx \int K^2(t)dt.$$

The excess of average squared error: Main results

Theorem 2, Benhenni, Girard, Louhichi (2021).

Suppose that Conditions on $(\epsilon_i)_i$ are satisfied. Then both

$$n(ASE(\hat{h}_{ML}) - ASE(\hat{h}_n)) \text{ and } n(ASE(\hat{h}_G) - ASE(\hat{h}_n))$$

converge in distribution to a $C\mathcal{X}_2(1)$ law, where C is the positive constant given by,

$$C = \frac{2\sigma^2}{5A}\left(\left(\int t^2K(t)dt\right)^2 \int_0^1 u^2(x)f''^2(x)dx \right.$$
$$\left. + \frac{2A}{B}\int_0^1 u^2(x)dx \int (K-G)^2(t)dt\right).$$

Application to ARCH(1) processes

Proposition

Let (ϵ_n) be a strictly stationary ARCH(1) process. Suppose that $\alpha^8 \prod_{i=1}^{8}(2i-1) < 1$. Then the asymptotic optimality, the asymptotic normality together with the conclusions of Theorem 2 are satisfied.

Simulation study for a trend plus ARCH(1) process

$K(x) = (15/8)(1 - 4x^2)^2 1_{[-.5,.5]}(x)$.

$$f(x) = (4x(1-x))^3,$$

and we use an equispaced design $x_i = i/n, i = 1, \cdots, n$,

$$\alpha \in \{0.01, 0.162, 0.577, 0.75, 0.9, 0.98\}$$

with a common value $\sigma = 0.32$. The first value of α nearly corresponds to i.i.d. normal observation noises (this setting will be referred to as the "quasi-iid-normal" case) and the last one generates noise sequences for which a strong serial correlation is always present when the sequence is squared. Recall that the moment of order 16 no longer exists as soon as α is slightly above 0.162, but the moment of order 4 still exists for $\alpha < \sqrt{1/3} \approx 0.57735$.

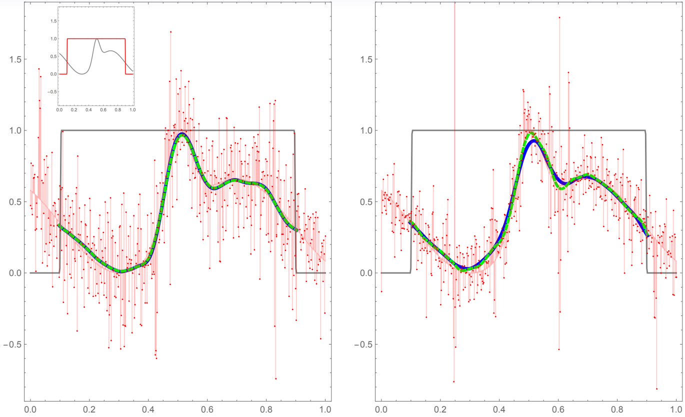
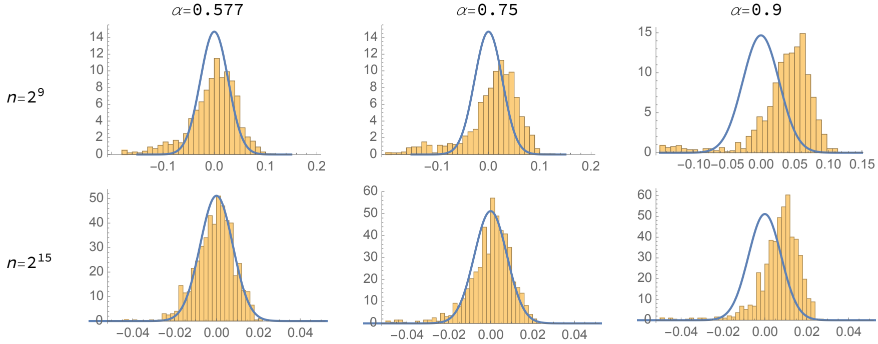


Figure 1 : $n = 2^9$. Each of these 2 panels displays one data set Y and the "smooth" deterministic trend $r(x)$. The 2 panels only differ by $\alpha = 0.577$ (left) and $\alpha = 0.9$ (right)


α=0.577
α=0.75
α=0.9
n=2^9
n=2^15
14
12
10
8
6
4
2
0
−0.1
0.0
0.1
0.2
15
5
−0.10−0.05 0.00 0.05 0.10 0.15
50
40
30
20
60
−0.04 −0.02 0.00 0.02 0.04

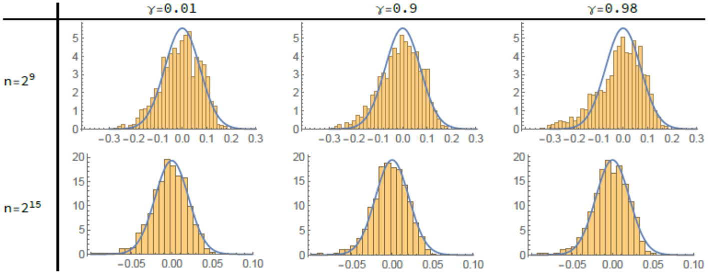
Log-normal volatility sequences

Corollary

Suppose that the volatility sequence $(\sigma_i)_{i\in\mathbb{N}}$ is defined for $i \in \mathbb{N}$, by $\sigma_i = \beta \exp\left(\sum_{j=0}^{\infty} \gamma^j \eta_{i-j}\right)$ with $|\gamma| < 1$, $\beta > 0$ and $(\eta_i)_{i\in\mathbb{Z}}$ is an i.i.d. centered sequence distributed as a Gaussian law with finite variance. Suppose also that Z_1 follows a standard Gaussian law. Then the process $(\epsilon_i)_{i\in\mathbb{N}}$ is a strictly stationary MDS, with finite all integer moments, strongly mixing with $\alpha_s = O(|\gamma|^{\frac{2}{3}s})$, and the asymptotic optimality, the asymptotic normality together with the conclusions of Theorem 2 hold.

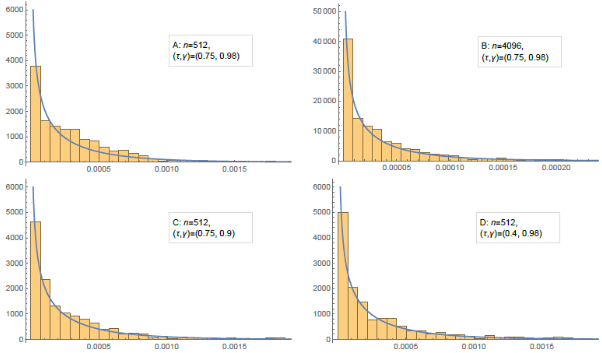
Simulations: Trend plus a log-normal SV process

$f(x) = (4x(1-x))^3$. These 6 panels only differ by n ($= 2^9$ in the top row and 2^{15} in the bottom row) and by γ varying in $\{0.01, 0.9, 0.98\}$. In each panel, the displayed normalized histogram is that of the 3000 replicates of $\hat{h}_M - \hat{h}_n$. The superposed blue curve is the normal distribution of $\hat{h}_M - \hat{h}_n$ as predicted by the asymptotic theory.



Simulations: Trend plus a log-normal SV process

Assessment of the simplified $\frac{6}{5n}\sigma^2\mathcal{X}_2(1)$ approximation of the ASE-excess. In each panel, the displayed normalized histogram is that of the 3000 replicates of $ASE(\hat{h}_M) - ASE(\hat{h}_n)$. The superposed blue curve is the $\frac{6}{5n}\sigma^2\mathcal{X}_2(1)$ density as suggested by the asymptotic theory. (A): $n = 2^9$ and $(\tau, \gamma) = (0.75, 0.98)$. (B): $n = 2^{12}$ and $(\tau, \gamma) = (0.75, 0.98)$. (C): $n = 2^9$ and $(\tau, \gamma) = (0.75, 0.9)$. (D): $n = 2^9$ and $(\tau, \gamma) = (0.4, 0.98)$.



We recall the following Marcinkiewicz-Zygmund type inequality which is a simple consequence of the Minkowski and the Burkholder inequalities (see Burkholder (1988)).

Theorem

Let $(\eta_i)_{i\geq 0}$ be a stationary centered sequence of martingale difference of finite pth moment with $p \geq 2$. Then there exists a positive constant c_p such that for any positive integer n,

$$\Big\| \sum_{i=1}^{n} \eta_i \Big\|_p^2 \leq c_p \sum_{i=1}^{n} \|\eta_i\|_p^2 .$$

An immediate consequence of Theorem 1 is the following corollary.

Corollary

Let $(\eta_i)_{i\geq 0}$ be a stationary sequence of martingale difference of finite pth moment with $p \geq 2$. Then there exists a positive constant c_p such that for any positive integer n,

$$\left\| \sum_{i=1}^{n} d_{i,n}\eta_i \right\|_p^2 \leq c_p \sum_{i=1}^{n} d_{i,n}^2,$$

and for any sequence of real numbers $(d_{i,n})_{1\leq i\leq n}$.

We also need the following proposition whose proof uses Theorem 1 above.

Proposition

Let $(\eta_i)_{i\geq 0}$ be a stationary sequence of martingale difference such that $\|\eta_i\|_{2p} < \infty$ for some $p \geq 2$. Then, there exists a positive constant c_p such that for any positive integer n,

$$\left\| \sum_{i=1}^{n} \sum_{j=1}^{i-1} b_{i,j,n} \eta_j \eta_i \right\|_p^2 \leq c_p \sum_{i=1}^{n} \sum_{j=1}^{i-1} b_{i,j,n}^2,$$

and for any sequence of real numbers $b_{i,j,n}$.

Let $X_i = \sum_{j=1}^{i-1} b_{i,j,n} \eta_j \eta_i$. The sequence $(X_i)_i$ is a martingale difference relative to the filtration $\sigma(\eta_j,\ j \leq i-1)$.

The following maximal inequality is also very needed in the proofs. Its proof needs some chaining arguments.

Lemma

Let $(\eta_i)_{i\geq 0}$ be a sequence of stationary martingale difference with $\|\eta_i\|_p < \infty$ for some $p \geq 2$. Let $(c_{i,n}(h))_{i,n,h}$ be a sequence of weights satisfying, for any $h, h' \in H_n = [an^{-1/5}, bn^{-1/5}]$,

$$|c_{i,n}(h) - c_{i,n}(h')| \leq cst\,|h - h'|.$$

and

$$\max_{i\leq n}\sup_{h\in H_n} |c_{i,n}(h)| \leq cst\, n^{-\alpha},\ \ \alpha > \frac{5p-2}{10(p-1)}.$$

Then,

$$\lim_{n\to\infty}\left\|\sup_{h\in H_n}\left|\sum_{i=1}^{n} c_{i,n}(h)\eta_i\right|\right\|_p = 0.$$

Lemma

Let $(\epsilon_j)_j$ be a sequence of random variables with finite fourth moment and such that,

$$\sup_i \sum_{j=1}^{\infty} |\mathrm{Cov}(\epsilon_i^2, \epsilon_j^2)| < \infty.$$

Let for $h \in H_n = [an^{-1/5}, bn^{-1/5}]$, $(d_{j,n}(h))_{1 \leq j \leq n}$ be a sequence of real numbers satisfying for any $1 \leq j \leq n$,

$$|d_{j,n}(h)| \leq \frac{cst}{n}, \;\; and, \;\; |d_{j,n}(h) - d_{j,n}(h')| \leq cst\, n^{-2/5}|h - h'|.$$

Then,

$$\lim_{n\to\infty} \left\| \sup_{h\in H_n} \left| \sum_{i=1}^{n} d_{i,n}(h)(\epsilon_i^2 - \mathbb{E}(\epsilon_i^2)) \right| \right\|_2 = 0.$$

Lemma

Let $(\eta_i)_{i\geq 0}$ be a stationary sequence of martingale difference random variables with finite moment of order $2p$, for some $p > 8$. Suppose that, for any $h, h' \in H_n$

$$|b_{i,j,n}(h)| \leq \frac{cst}{n} \mathbb{1}_{|i-j|\leq 2nh},$$

$$|b_{i,j,n}(h) - b_{i,j,n}(h')| \leq cst\, n^{-4/5} |h - h'| \mathbb{1}_{|i-j|\leq 2n \max(h,h')}.$$

Then,

$$\lim_{n\to\infty} \left\| \sup_{h\in H_n} \left| \sum_{i=1}^{n} \sum_{j=1}^{i-1} b_{i,j,n}(h) \eta_j \eta_i \right| \right\|_p = 0.$$

CLT

Recall that $K - G$ is an even function, $[-1,1]$-supported, that the window h_n is a positive sequence satisfying

$$\lim_{n\to\infty} h_n = 0, \quad \lim_{n\to\infty} nh_n = \infty.$$

Define, for $i = 1, \cdots, n$, $x_i = \frac{i}{n}$ and, for a positive constant C_K depending only on K,

$$a_{i,n}(h_n) = C_K \frac{h_n}{n} f''(x_i)u(x_i)$$
$$b_{i,j}(h_n) = \frac{1}{n^2 h_n^2}(K-G)(\frac{x_i - x_j}{h_n})$$
$$\tilde{b}_{i,j} = b_{i,j}(h_n)(u(x_i) + u(x_j)).$$

Let $(\epsilon_i)_{i\geq 0}$ be a centered sequence of stationary MD random variables with finite second moment σ^2. Let

$$Y_{i,n}(h_n) = a_{i,n}(h_n)\epsilon_i + \sum_{j=1}^{i-1} \tilde{b}_{i,j}\epsilon_i\epsilon_j, \qquad (2)$$

CLT: Control of the variance

Proposition

Suppose that there exists a positive decreasing function Φ defined on $\mathbb{R}^+$ satisfying

$$\sum_{s=1}^{\infty} s^2 \Phi(s) < \infty,$$

and for any $1 \leq i_1 \leq i_2 < i_3 \leq i_4 \leq i_5 \leq n$ such that $i_3 - i_2 \geq \max(i_2 - i_1, i_4 - i_3, i_5 - i_4)$

$$|\mathrm{Cov}(\epsilon_{i_1}\epsilon_{i_2}, \epsilon_{i_3}\epsilon_{i_4})| \leq \Phi(i_3 - i_2)$$
$$|\mathrm{Cov}(\epsilon_{i_2}, \epsilon_{i_3}\epsilon_{i_4}\epsilon_{i_5})| \leq \Phi(i_3 - i_2).$$

Then

$$\mathrm{Var}\left(\sum_{i=1}^{n} Y_{i,n}(h_n)\right) = \frac{h_n^2\sigma^2}{n} C_K^2 \int u^2(x) f''^2(x)dx$$
$$+\frac{4\sigma^4}{n^2 h_n^3}\int_0^1 u^2(x)dx \int_0^1 (K-G)^2(u)du$$
$$+o(\frac{1}{n^2 h_n^3} + \frac{h_n^2}{n}).$$

Proposition

Let $(\epsilon_i)_{i\geq 0}$ be a stationary sequence of centered martingale difference random variables relative to the filtration $\mathcal{F}_i = \sigma(\epsilon_1, \cdots, \epsilon_i)$. Suppose that $\mathbb{E}(\epsilon_1^8) < \infty$. Suppose, moreover, that there exists a positive decreasing function Φ defined on $\mathbb{R}^+$ satisfying

$$\sum_{s=1}^{\infty} s^4 \Phi(s) < \infty,$$

and for any positive integer $q \leq 6$, $1 \leq i_1 \leq \cdots \leq i_k < i_{k+1} \leq i_q \leq n$ such that $i_{k+1} - i_k \geq \max_{1\leq l \leq k}(i_{l+1} - i_l)$

$$|\mathrm{Cov}(\epsilon_{i_1} \cdots \epsilon_{i_k}, \epsilon_{i_{k+1}} \cdots \epsilon_{i_q})| \leq \Phi(i_{k+1} - i_k).$$

Let $Y_{in}(h_n)$ be as defined in (2) with $h_n = cn^{-1/5}$.

Then

$$n^{7/10} \sum_{i=1}^{n} Y_{i,n}(h_n) \Longrightarrow \mathcal{N}(0, V),$$

where $\Longrightarrow$ denotes convergence in distribution when n tends to infinity, the variance V is defined by,

$$V = c^2 C_K^2 \sigma^2 \int_0^1 u^2(x) f''^2(x) dx$$
$$+ \frac{4}{c^3} \sigma^4 \int_0^1 u^2(x) dx \int_0^1 (K - G)^2(u) du,$$

and $\sigma^2 = \mathbb{E}(\epsilon_1^2)$.

Main tools for the control of the excess of errors

$$
\begin{aligned}
& n(ASE(\hat{h}) - ASE(\hat{h}_n)) \\
&= n(\hat{h} - \hat{h}_n)ASE_n'(\hat{h}_n) + \frac{n}{2}(\hat{h} - \hat{h}_n)^2 ASE''(h^*) \\
&= \frac{n}{2}(\hat{h} - \hat{h}_n)^2 \mathbb{E}(ASE''(h_n^*)) \\
&+ \frac{n}{2}(\hat{h} - \hat{h}_n)^2 \left(ASE''(h^*) - \mathbb{E}(ASE''(h_n^*)) \right),
\end{aligned}
$$

$$
\Sigma^{-1} n^{3/10}(\hat{h} - \hat{h}_n) \Longrightarrow \mathcal{N}(0,1),
$$

$$
\lim_{n\to\infty} n^{2/5}\mathbb{E}(ASE''(h_n^*)) = 5\sigma^{4/5} B^{2/5} A^{3/5},
$$

- $\left| \frac{\hat{h}}{h_n^*} - 1 \right| \to 0$ a.s.
- $\sup_{h \in H_n} \left| ASE''(h) - \mathbb{E}(ASE''(h)) \right|$,

  $\sup_{|h_1 - h_2| \le a_n} \left| \mathbb{E}(ASE''(h_1)) - \mathbb{E}(ASE''(h_2)) \right|$

Some References

- Girard, D.A. (1998). Ann. Stat., 26, 315-334.
- Doukhan, P. and Louhichi, S. (1999). Stochastic Process. Appl., 84, 2, 313-342.
- Han, Ch. and Gu, Ch. (2008). Sankhya: The Indian Journal of Statistics, Series A, 70, 1, 38-72.
- Härdle, W., Hall, P. and Marron, J.S. (1988). Journal of the American Statistical Association 83, 86-95.
- Hall, P., Lahiri, S. N. and Polzehl, J. (1995). Ann. Statist. 23-6, 1921-1936.
- Mallows, C. L. (1973). Technometrics 15, 661-675.
- Rice, J. (1984). Annals of Statistics 12, 1215-1230.
- Rio, E. (1993). Ann. Inst. H. Poincaré Probab. Statist., 29 587-597.

THANK YOU FOR YOUR ATTENTION