ADAPTIVE SAMPLING SCHEMES FOR DENSITY ESTIMATION

Delphine Blanke

January 2005

- Estimation with known parameters : r_0, γ_0
- Adaptive sampling scheme : r_0 or γ_0 unknown
- Double adaptive sampling scheme : r_0, γ_0 both unknown

DENSITY - n I.I.D. OBS.

$X_1, \ldots, X_n$, i.i.d., $\mathbb{R}^d$-valued, $X_i \sim f$.

Regularity condition, $\mathcal{C}^{(r_0)}$

$$\begin{cases} f \text{ is } r_0 - \text{times differentiable with either} \\ f^{(r_0)} \text{ a bounded continuous function or for} \\ j_1 + \cdots + j_d = r_0 \ \text{ and } (\ell, \lambda) \in]0, +\infty[\times]0, 1] \ : \\ \qquad \left| \frac{\partial^{(r_0)} f}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}}(y) - \frac{\partial^{(r_0)} f}{\partial y_1^{j_1} \ldots \partial y_d^{j_d}}(z) \right| \leq \ell \|y - z\|^{\lambda}. \end{cases}$$

Kernel estimator

$$f_n(x) = \frac{1}{n h_n^d(r_0)} \sum_{i=1}^{n} K_{(r_0)} \left(\frac{x - X_i}{h_n(r_0)} \right), \quad x \in \mathbb{R}^d$$

for bandwidth $h_n(r_0)$, kernel $K_{(r_0)}$ with compact support and belonging to $\mathcal{C}_K^{(r_0)}$.

Mean-square convergence $h_n(r_0) \sim (1/n)^{\frac{1}{2r_0+d}}$

$$n^{\frac{2r_0}{2r_0+d}} \mathrm{E}\left(f_n(x) - f(x)\right)^2 \xrightarrow[n\to\infty]{} f(x) \int K^2_{(r_0)}(v)\,\mathrm{d}v + b^2_{r_0}(x) \qquad (*)$$

with $b_{r_0}(x) =$

$$\sum_{j_1+\cdots+j_d=r_0} \frac{1}{j_1!\ldots j_d!} \frac{\partial^{(r_0)} f(x)}{\partial x_1^{j_1} \ldots \partial x_d^{j_d}} \int u_1^{j_1} \ldots u_d^{j_d} K(u)\,\mathrm{d}u.$$

Motivation

- Get the result $(*)$ for a continuous-time process sampled at n instant times, $t_{i,n} = \delta_n$.
- High rate sampling ($\delta_n \to 0$, $n\delta_n \to \infty$) should be adapted to the feature of the underlying sample path.

DENSITY - SAMPLING

$\{X_t,\, t \in \mathbb{R}\}$: (X_t) $\mathbb{R}^d$-valued, defined on $(\Omega, \mathcal{A}, \mathcal{P})$, measurable, with same density f.

Sampling times : $0 < t_{1,n} < \cdots < t_{n,n}$ such that

$$t_{k,n} - t_{k-1,n} = \delta_n, \qquad k = 2, \ldots, n$$

with δ_n satisfying to $\delta_n \to 0$, $n\delta_n \to +\infty$.

Kernel estimator

$$\hat{f}_n(x) = \frac{1}{nh_n^d(r_0)} \sum_{i=1}^{n} K_{(r_0)} \left(\frac{x - X_{t_{i,n}}}{h_n(r_0)} \right).$$

Some references, $\delta_n \equiv \delta > 0$: Masry (83), Prakasa Rao (88), Wu (97), Vilar & Vilar (00)...

$\delta_n \to 0$: Bosq (97, 98), Bosq & Cheze-Payaud (99), Leblanc (95, 97), Comte & Merlevède (02) ...

Question : How to choose δ_n minimal ?

$\rightsquigarrow$ Minimal total time of experiment $T_n = n\delta_n$.

Assumptions 1 (A1)

(i) f *bounded, continuous at* x *;*

(ii) $f_{(X_s,X_t)} = f_{(X_0,X_{t-s})}$ *for* $t > s$ *;*

(iii) $\exists\, u_1 \geq u_0 > 0 :\ \forall\, u \in [u_0, +\infty[$,
$\|g_u\|_\infty \leq \pi(u)$ *with* π *bounded, integrable*

and $\searrow$ *over* $]u_1, +\infty[$,

(iv) $\exists\, \gamma_0 > 0 :$ $f_{(X_0,X_u)}(y,z) \leq M(y,z)u^{-\gamma_0}$,

for $(y,z,u) \in \mathcal{V}_x^2 \times]0, u_0[$, *with* $M(.,.)$ *continuous at* (x,x).

Remark : If $Y_u := \left(\frac{X_u^{(1)} - X_0^{(1)}}{u^{\gamma_1}}, \ldots, \frac{X_u^{(d)} - X_0^{(d)}}{u^{\gamma_d}} \right)$

with $X_t^{(i)}$ i-th component of $\left(X_t^{(1)}, \ldots, X_t^{(d)}\right)$, $0 < \gamma_i \leq 1$ for $i = 1, \ldots, d$; then A1(iv) can be replaced by $f_{(X_0,Y_u)}(y, \frac{z-y}{u^\gamma}) \leq M(y,z)$, with

$\gamma_0 = \sum_{i=1}^{d} \gamma_i$ and $\frac{z-y}{u^\gamma} := \left(\frac{z_1 - y_1}{u^{\gamma_1}}, \ldots, \frac{z_d - y_d}{u^{\gamma_d}}\right)$.

- Case $\gamma_0 = \frac{1}{2}$. Homogeneous diffusions : under regularity conditions, solutions of

$$\mathrm{d}X_t = m(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t,\ t \geq 0$$

are strictly stationary and satisfy A1-(iii), A1-(iv) (with $\gamma_0 = 1/2$), see e.g. Leblanc (1997), Veretennikov (1999), Kutoyants (2003).

- Case $\gamma_0 = 1$. For example, real mean-square differentiable Gaussian processes. But also 2-dimensional homogeneous diffusion processes with independent components $X_t^{(1)}$, $X_t^{(2)}$.

- Case $\gamma_0 > 1$: e.g. d-dimensional diffusion processes, $d \geq 3$ independent components, with $\gamma_0 = \sum_{i=1}^{d} \gamma_i$.

MEAN-SQUARE CONVERGENCE

Theorem 1 *(with Pumo, 2003)*

Under Assumption A1, $f \in \mathcal{C}^{(r_0)}$, $h_n = n^{-\frac{1}{2r_0+d}}$ *:*

$$n^{\frac{2r_0}{2r_0+d}} E(\hat{f}_n(x) - f(x))^2 \to f(x) \int K^2_{(r_0)}(u)\, du + b^2_{r_0}(x)$$

with δ_n *such that* $\delta_n/\delta_n^*(\gamma_0) \to \infty$ *where*

$$\begin{cases} \delta_n^*(\gamma_0) = h_n^d & \textit{if} \quad \gamma_0 < 1, \\ \delta_n^*(\gamma_0) = h_n^d \ln(1/h_n) & \textit{if} \quad \gamma_0 = 1, \\ \delta_n^*(\gamma_0) = h_n^{d/\gamma_0} & \textit{if} \quad \gamma_0 > 1. \end{cases}$$

'Optimality ?' : Optimality in the following sense : if $T_n = n\delta_n^*(\gamma_0)$, then for $\gamma_0 < 1$, $\gamma_0 = 1$ and $\gamma_0 > 1$ the corresponding m.s. rates are T_n^{-1}, $\frac{\ln T_n}{T_n}$ and $T_n^{-\frac{2\gamma_0 r_0}{2\gamma_0 r_0 + d(\gamma_0 - 1)}}$.

$\rightsquigarrow$ Castellana & Leadbetter (86), Bosq (97, 98), Bosq & Davydov (98), Davydov (01), Kutoyants (97, 99, 03), Bl. & Bosq (97, 00), Sköld & Hössjer (99)...

ALMOST SURE CONVERGENCE

Theorem 2 *X geometrically strong mixing proc.*

(a) if $\delta_n \equiv \delta$, conditions A1(i)-(iii) and $\frac{nh_n^d}{(\ln n)^3} \to \infty$ imply a.s.

$$\limsup_{n\to+\infty} \sqrt{\frac{nh_n^d}{\ln n}} \left| \hat{f}_n(x) - \mathrm{E}\,\hat{f}_n(x) \right| \leq 2^{\frac{3}{2}} f^{\frac{1}{2}}(x) \, \|K\|_2 \qquad (*)$$

(b) $(*)$ also true if moreover A1(iv) is satisfied and $h_n \to 0$, $\delta_n \to 0$ such that $\frac{nh_n^d \delta_n^2}{(\ln n)^3} \to \infty$ and $\delta_n \geq \delta_n^*(\gamma_0)$ where

$$\delta_n^*(\gamma_0) = \begin{cases} h_n^d & \text{if } \gamma_0 < 1, \\ h_n^d \ln(1/h_n) & \text{if } \gamma_0 = 1, \\ h_n^{d/\gamma_0} & \text{if } \gamma_0 > 1. \end{cases}$$

Remark : $f \in C^{r_0}$ and $r_0 > \frac{d}{\max(1,\gamma_0)}$ $\rightsquigarrow$ rate $\left(\frac{\ln n}{n}\right)^{\frac{r_0}{2r_0+d}}$ for $h_n \sim \left(\frac{\ln n}{n}\right)^{\frac{1}{2r_0+d}}$ and for all $\delta_n \geq \delta_n^*(\gamma_0)$. If $r_0 \leq \frac{d}{\max(1,\gamma_0)}$, a suitable choice is $\delta_n = \frac{(\ln n)^{3/2}}{(nh_n^d)^{1/2}} \ln_p n$ with $\ln_p(n) = \underbrace{\ln \cdots \ln}_{p-\text{times}} n$, $p \geq 2$.

Some simulations

Numerical implementation in the Gaussian case (with Pumo, 03).

- Criteria

$$ISE(\delta) = \frac{1}{N}\sum_{j=1}^{N}\int\left(\hat{f}_{n,\delta,j}(x) - f(x)\right)^2 dx$$

with $t_{i+1,n} - t_{i,n} = \delta$, $f \sim \mathcal{N}(0,1)$, $n = 105$, $N = 50$ and $\hat{f}_{n,\delta,j}$ estimator for the j-th simulated sample path.

- Ornstein-Uhlenbeck ($\gamma_0 = 1/2$)

$dX_t = -X_t dt + \sqrt{2} dW_t$

and Wong process ($\gamma_0 = 1$)

$$X_t = \sqrt{3}\exp(-\sqrt{3}t)\int_0^{\exp(2t/\sqrt{3})} W_s\, ds,$$

both simulated at times $\tau_{m+1} - \tau_m = 0.02$.

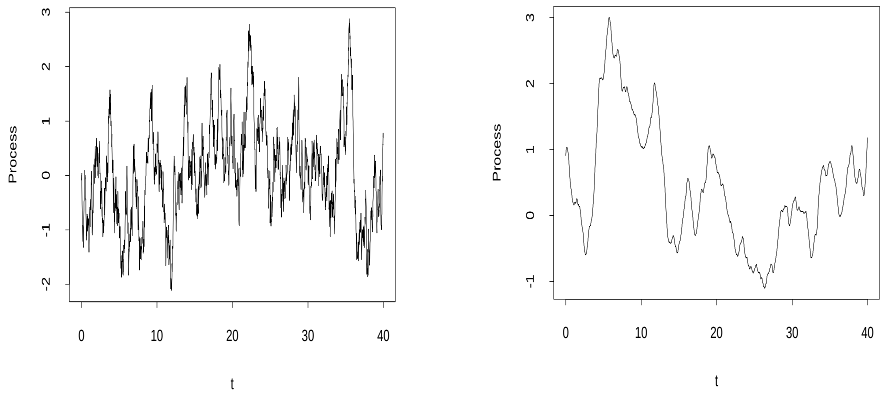


Figure 1: O.U. (left) and Wong (right) evaluated at $\tau_i = 0.02 * i$.

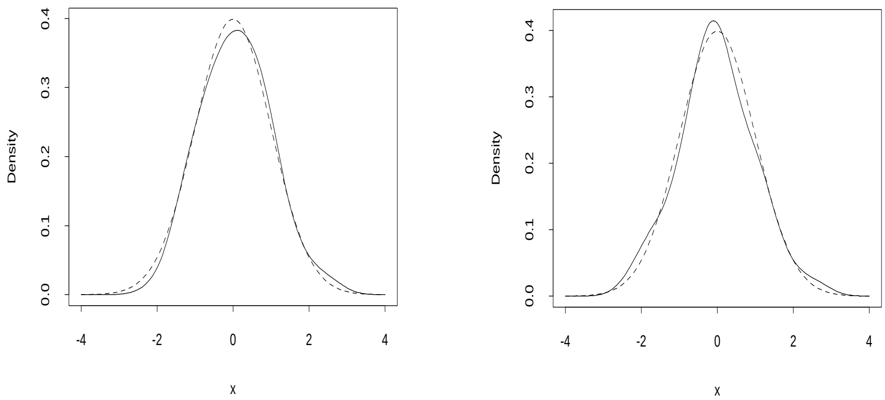


Figure 2: Estimated density (dash) $\mathcal{N}(0, 1)$, $n = 105$.
O.U. (left) for $\delta_n = 0.4$ and Wong (right) for $\delta_n = 1.83$.

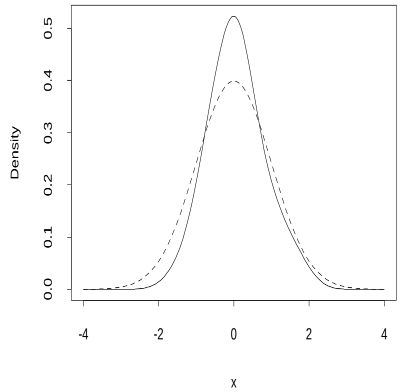
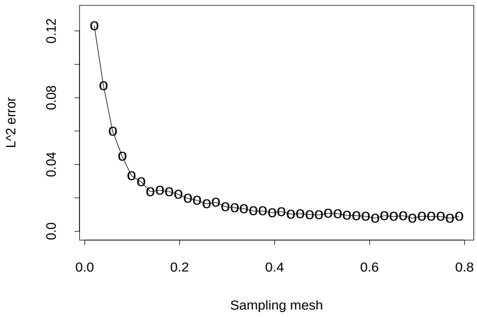


Figure 3: Estimated density (dash) $\mathcal{N}(0,\, 1)$, $n = 105$.
Wong with $\delta_n = 0.4$.



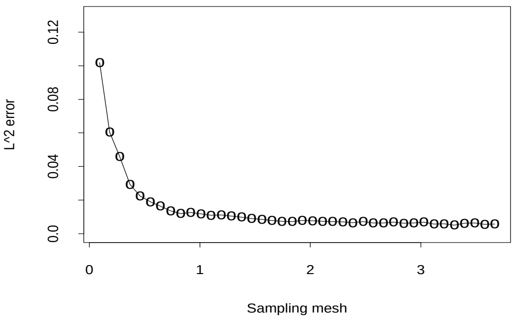


Figure 4: $ISE(\delta)$ for O.U. (left) and Wong (right).

$$ISE(\delta) = \frac{1}{50} \sum_{j=1}^{50} \int (\widehat{f}_{105,\delta,j}(x) - f(x))^2 \, \mathrm{d}x$$

ADAPTIVE SAMPLING SCHEME

(r_0 unknown, γ_0 known)

Some references on adaptive works

• Lepski (90) : white noise model, Hölder class.

• Efromovich (85) : i.i.d. case , density on [0,1], Sobolev class, L^2.

• Golubev (92) : i.i.d. case, Sobolev class, L^2.

• Butucea (00, 01) : i.i.d. case , Sobolev class, L^p.

• Comte & Merlevède (02) : strong mixing process, $\gamma_0 < 1$, Besov space.

Our framework A device to be calibrated during a learning period where different choices of δ_n could be tested. Construction of an adaptive estimator (relatively to r_0 and for known γ_0) converging over the smallest possible learning period.

Grid of candidates for r_0 :

$\Delta_n = \{1, 2, \ldots, r_n\}$

$\rightsquigarrow$ $r_0^* = \max\{r_1 \in \Delta_n : \forall\, r_2 \in \Delta_n, r_2 \leq r_1,$
$(\frac{n}{\ln n})^{\frac{r_2}{2r_2+d}} |\widehat{f}_{r_2,\gamma_0}(x) - \widehat{f}_{r_1,\gamma_0}(x)| \leq \widehat{\eta}(r_2, \gamma_0)\}$
where $\widehat{\eta}(r, \gamma_0)$ is some random quantity to be defined,

$$\widehat{f}_{r,\gamma_0}(x) := \frac{1}{nh_n^d(r)} \sum_{i=1}^{n} K_{(r)} \left(\frac{x - X_{i\delta_h(r,\gamma_0)}}{h_n(r)} \right),$$

with $h_n(r) = c \left(\frac{\ln n}{n} \right)^{\frac{1}{2r+d}}$, $K_{(r)} \in \mathcal{C}_K^{(r)}$ and $\delta_h(r, \gamma)$ such that

$$\begin{cases} \delta & \text{if} \quad \gamma = \gamma_\infty \\ h_n^d(r) \ln_p(n) & \text{if} \quad \gamma < 1,\ r > d \\ h_n^d(r) \ln \ 1/h_n(r) \ \ln_p(n) & \text{if} \quad \gamma = 1,\ r > d \\ h_n^{d/\gamma}(r) \ln_p(n) & \text{if} \quad \gamma > 1,\ r > d/\gamma \\ \frac{(\ln n)^3}{nh_n^d(r)} \ ^{1/2} \ln_p(n) & \text{if} \quad r \leq d/\max(\gamma, 1). \end{cases}$$

Note that $\gamma_0 = \gamma_\infty$ represents the fixed design $\delta_n \equiv \delta$.

$\rightsquigarrow$ Adaptive estimator :

$$\widehat{f}_{r_0^*,\gamma_0}(x) = \frac{1}{nh_n^d(r_0^*)} \sum_{i=1}^{n} K_{(r_0^*)} \left(\frac{x - X_{i\delta_h(r_0^*,\gamma_0)}}{h_n(r_0^*)} \right)$$

with $h_n(r_0^*) = c\left(\frac{\ln n}{n}\right)^{\frac{1}{2r_0^*+d}}$. Note that by setting $\gamma_0 = \gamma_\infty$ one obtains adaptive estimation of f relatively to r_0 in the case of the fixed design $\delta_n \equiv \delta$.

Now asymptotic convergence of $\widehat{f}_{r_0^*,\gamma_0}(x)$ depends on $\widehat{\eta}(r,\gamma_0) =$

$$a\left(2^{3/2}c^{-d/2}\sqrt{\widetilde{f}_{r,\gamma_0}(x)}\,\|K_{(r)}\|_2 + c^r\left|\widetilde{b}_{r,\gamma_0}(x)\right|\right)$$

with $a > 2$ and $\widetilde{f}_{r,\gamma_0}(x)$, $\widetilde{b}_{r,\gamma_0}(x)$ preliminary estimators of $f(x)$, $b_r(x)$ based on $X_{\delta_n(\gamma_0)}, \ldots, X_{n\delta_n(\gamma_0)}$.

Lemma 1 *Under A1, X G.S.M., $f \in \mathcal{C}^{(r_0)}$ then for all $r_k \le r_0$, $\widehat{\eta}(r_k,\gamma_0) \xrightarrow[n\to\infty]{} a\,C_1(r_k)$ a.s. with $C_1(r_k) = 2^{3/2}c^{-d/2}f^{1/2}(x)\,\|K\|_2 + c^{r_k}\,|b_{r_k}(x)|$ and $r_n (= \sharp\Delta_n) \to \infty$ such that $r_n = \mathcal{O}\left(\frac{\ln n}{(\ln\ln n)^{\nu_2}}\right)$, $\nu_2 > 1$.*

Finally, we may state our main result :

Theorem 3 *If conditions of Lemma 1 are fulfilled, one obtains a.s.*

$$\limsup_{n\to\infty}\left(\frac{n}{\ln n}\right)^{\frac{r_0}{2r_0+d}}\left|\widehat{f}_{r_0^*,\gamma_0}(x)-f(x)\right|\le(a+1)C_1(r_0)$$

Remarks

- Construction of $\widehat{f}_{r_0^*,\gamma_0}$ requires the sequence of observations
  $X^{r,n}=\left(X_{i\delta(r,\gamma_0)},\ i=1,\dots,n\right),\, r=\mathcal{O}(r_n).$
  Since $r_n=o\big(\ln n\big)$, the loss of rate is at most logarithmic.

- Theorem 3 remains true if one works only with an upper bound for γ_0, say $\overline{\gamma_0}$, since then $\delta_n(\overline{\gamma_0})\ge\delta_n^*(\gamma_0)$, the minimal sampling rate.
  Recall that in general, $\gamma_0=\sum_{i=1}^{d}\gamma_i,\ \gamma_i\in]0,1]$
  $\rightsquigarrow\overline{\gamma_0}=d.$

- The case r_0 known but γ_0 unknown can be similarly handled (Bl., 2003).

A DOUBLE ADAPTIVE SAMPLING SCHEME

(r_0, γ_0 both unknown)

Framework : Some device to be calibrated during a learning period where various sampling rates δ_n can be adjusted (including the fixed one $\delta_n \equiv \delta$). The goal is to keep the minimal one, say δ_n^*, satisfying to $\delta_n^* \geq \delta_n^*(\gamma_0)$ when r_0, γ_0 are both unknown.

Grid for γ_0 :

$$\Gamma_n = \{\widetilde{\gamma}_0, \gamma_{1,n}, \gamma_{2,n}, \ldots, \gamma_{N_n,n}, \gamma_\infty\}$$

with $0 < \widetilde{\gamma}_0 < 1$, $\gamma_{1,n} = 1$, $\gamma_{j+1,n} - \gamma_{j,n} = \tau_n$,

γ_∞ corresponding to $\delta_n \equiv \delta$ and

$$\begin{cases} N_n \tau_n & = & o(\ln n), \\ \text{and } \tau \geq \tau_n & \geq & \frac{(\ln \ln n)^{\nu_1}}{\ln n}, \quad (\nu_1 > 1). \end{cases}$$

Grid for r_0 : $\Delta_n = \{1, 2, \ldots, r_n\}$ with $r_n = \mathcal{O}\left(\frac{\ln n}{(\ln \ln n)^{\nu_2}}\right)$, $\nu_2 > 1$.

Procedure in 2 steps :

- For $\delta_n \equiv \delta$, look at a candidate r_0^* for r_0 in Δ_n :

$$r_0^* = \max\{r_1 \in \Delta_n : \forall\, r_2 \in \Delta_n, r_2 \le r_1,\ \left(\frac{n}{\ln n}\right)^{\frac{r_2}{2r_2+d}} |\widehat{f}_{r_2,\gamma_\infty}(x) - \widehat{f}_{r_1,\gamma_\infty}(x)| \le \widehat{\eta}(r_2,\gamma_\infty)\}$$

with $\widehat{\eta}(r,\gamma_\infty)$ defined as before.

- Look at a candidate γ_0^* in Γ_n with the help of r_0^* :

$$\gamma_0^* = \min\{\gamma_1 \in \Gamma_n :\ \forall\, \gamma_2 \in \Gamma_n,\ \gamma_2 \ge \gamma_1,\ \left(\frac{n}{\ln n}\right)^{\frac{r_0^*}{2r_0^*+d}} |\widehat{f}_{r_0^*,\gamma_2}(x) - \widehat{f}_{r_0^*,\gamma_1}(x)| \le \zeta\}$$

with $\zeta > 0$.

The adaptive estimator

$$\widehat{f}_{r_0^*,\gamma_0^*}(x) = \frac{1}{nh_n^d(r_0^*)} \sum_{i=1}^{n} K_{(r_0^*)}\left(\frac{x - X_{i\delta_h(r_0^*,\gamma_0^*)}}{h_n(r_0^*)}\right)$$

with $h_n(r) = c\left(\frac{\ln n}{n}\right)^{\frac{1}{2r+d}}$ and $\delta_h(r,\gamma)$ defined as follows :

$$\begin{cases} \delta & \text{if} \quad \gamma = \gamma_\infty \\ h_n^d(r)\ln_p(n) & \text{if} \quad \gamma < 1,\ r > d \\ h_n^d(r)\ln(1/h_n(r))\ln_p(n) & \text{if} \quad \gamma = 1,\ r > d \\ h_n^{d/\gamma}(r)\ln_p(n) & \text{if} \quad \gamma > 1,\ r > d/\gamma \\ \sqrt{\frac{(\ln n)^3}{nh_n^d(r)}}\ln_p(n) & \text{if} \quad r \le d/\max(\gamma,1). \end{cases}$$

Theorem 4 *Under A1 and if X is a geometrically strongly mixing proc., for $f \in \mathcal{C}^{r_0}$ one obtains a.s.*

$$\limsup_{n\to\infty}\left(\frac{n}{\ln n}\right)^{\frac{r_0}{2r_0+d}}\left|\widehat{f}_{r_0^*,\gamma_0^*}(x)-f(x)\right| \le (a+1)C_1+\zeta$$

Remarks

- Observations needed for estimation $X^{s,r,n}=\left(X_{i\delta(r,\gamma_s)}, i=1,\dots,n\right)$, $s=0,1,\dots,N_n+1$, $r=1,\dots,r_n$ with $N_n=o\left((\ln n)^2\right)$ and $r_n=o\left(\ln n\right)$ $\rightsquigarrow$ only a logarithmic loss in relation to an estimator using the whole $N_n \times r_n \times n$ observations.
- $\tau_{j,n}\equiv\tau$, $N_n\equiv N$ is a suitable choice $\rightsquigarrow$ the numerical implementation can be fast.
- γ_0^* quite bad estimator of γ_0 !